Case 4:22-cv-01490-JST Document 99-2 Filed 06/16/23 Page 1 of 151

### **EXHIBIT A**

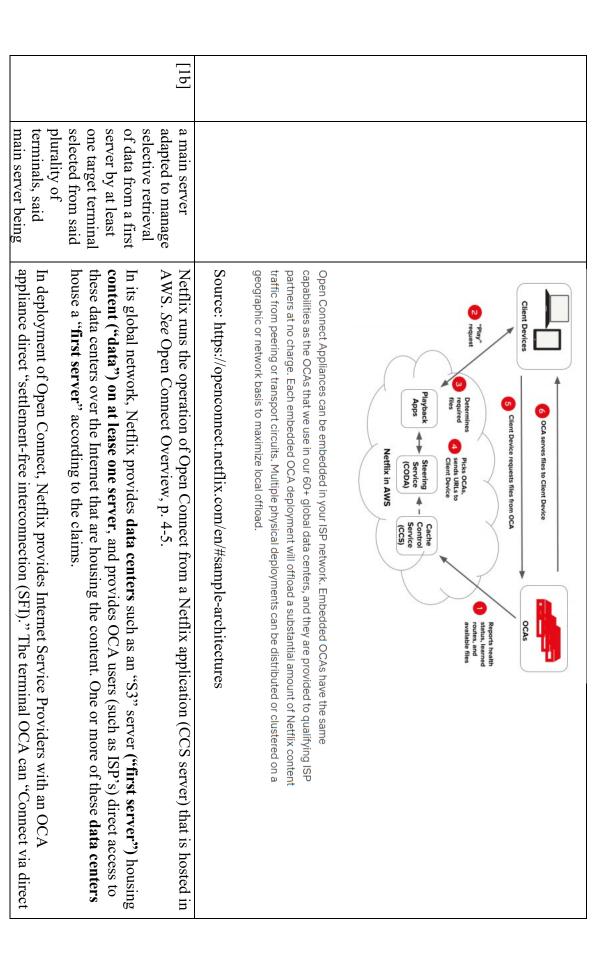
# PLAINTIFF'S SUPPLEMENTAL INFRINGEMENT CONTENTIONS

June 13, 2023

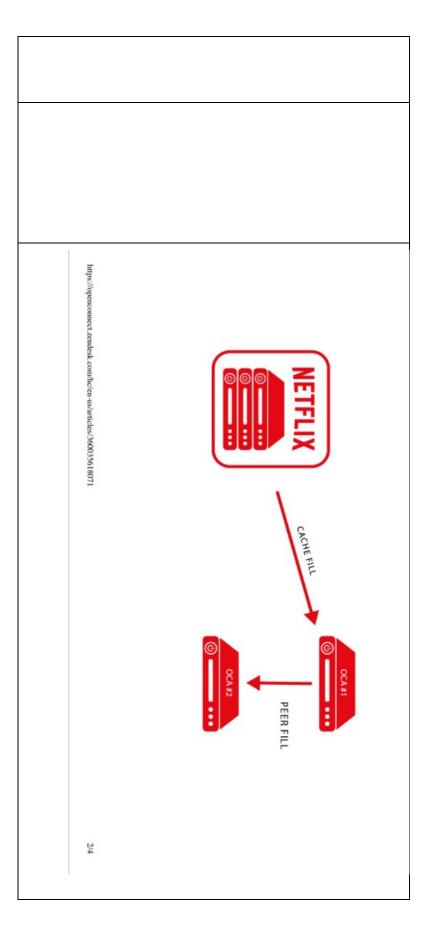
Claim Chart for U.S. Patent 8,495,167 - "Data communications networks, systems, methods and apparatus"

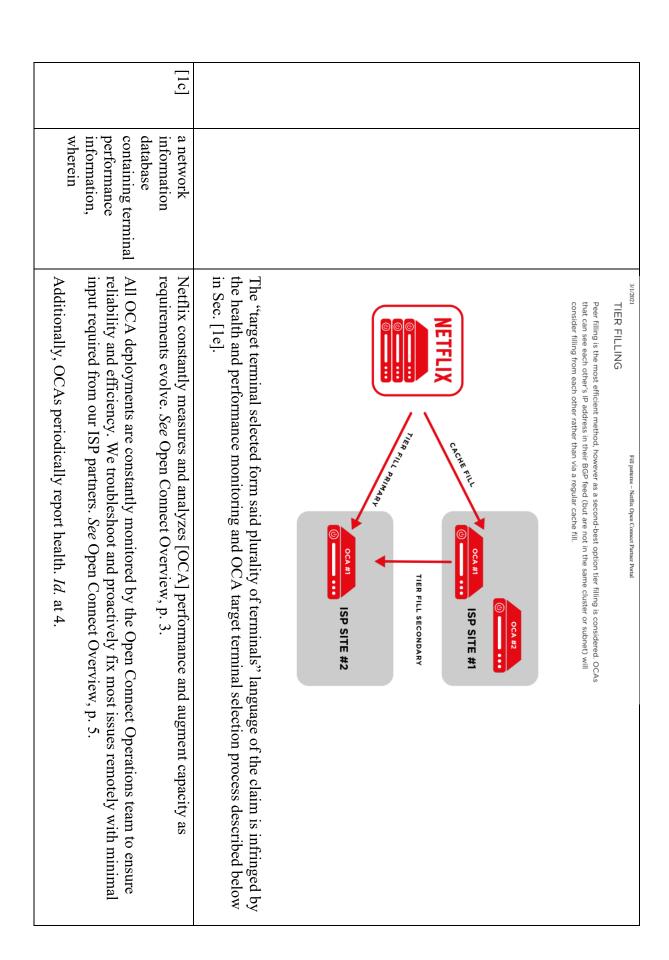
US Patent 8,495,167

Priority Date: Jul 30, 2002 Filing Date: Jul. 30, 2002



					d fi
					distinct from said first server; and
OCAs in a cluster and on the same subnet can attempt peer filling from each other. There is also Tier filling where if in different ISP sites. <a href="deploymentguide.pdf">deploymentguide.pdf</a> (netflix.com)	The following diagram also illustrates access from a target terminal (OCA $\#1$ , OCA $\#2$ ) to a Netflix first server in "our data centers."	Welcome to Open Connect, p. 3. Dkt 39 at p. 48.	ISPs who do not currently participate in pubic peering might want to consider that a single IX port can support multiple peering sessions, providing <b>direct access to various content</b> , cloud, and network providers.	"Netflix has the ability to interconnect at a number of global data center facilities and public Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any network at IXP locations where we are mutually present and we consider private interconnection as appropriate."	Private Network Interconnect (PNI) or IXP-based SFI peering to Netflix Open Connect Appliances in our data centers."





## Monitoring, Maintenance, and Updates

performance degrades to the point where a server is no longer functioning in the range of our quality standards, we simply replace it - at no cost to our partners. status and performance, we provide a Partner Portal where they can do so. If hardware required from our ISP partners. If partners wish to monitor their own embedded OCAs' efficiency. We troubleshoot and proactively fix most issues remotely with minimal input constantly monitored by the Open Connect Operations team to ensure reliability and All of our OCA deployments, whether in IXPs or embedded in ISP networks, are

Netflix wrote: of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a>, and in which terminals are selected to act as relay servers for a particular data transfers on the basis The response from the control plane in AWS is a ranked list of potential download locations, a.k.a The main server on CCS is further adapted to monitor response times of terminals in the network

fill sources, for each title. The determination of the list takes into consideration several high-level

Title (content) availability — Does the fill source have the requested title stored?

# Fill health — Can the fill source take on additional fill traffic?

A calculated route cost — Described in the next section

A fill escalation policy defines:

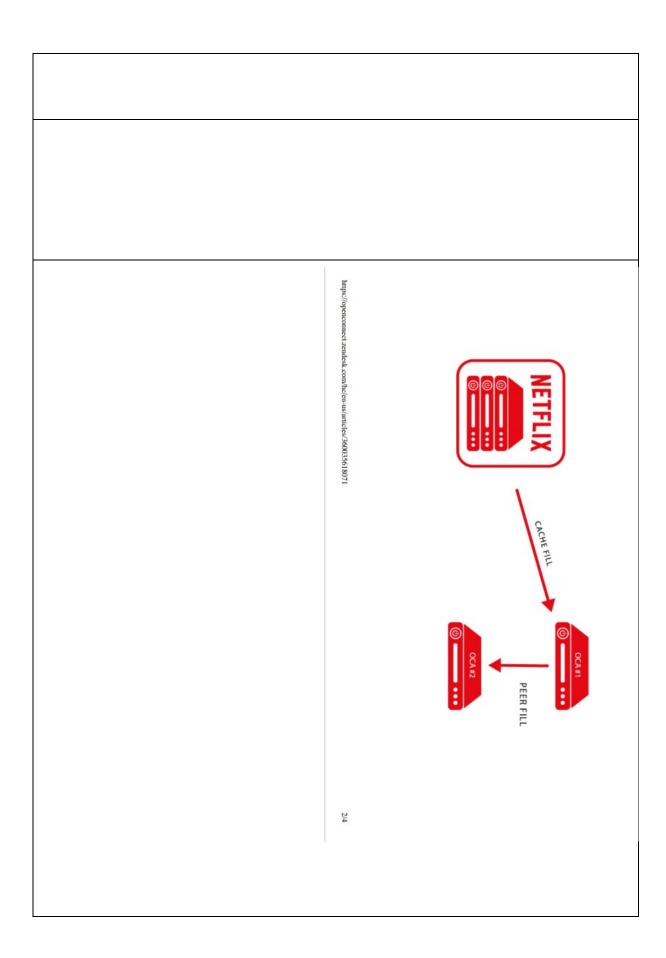
- How many hops away an OCA can go to download content, and how long it should wait before doing so
- above), and how long it should wait before doing so Whether the OCA can go to the entire Open Connect network (beyond the hops defined

2

Whether the OCA can go to S3, and how long it should wait before doing so

(Emphasis added.)

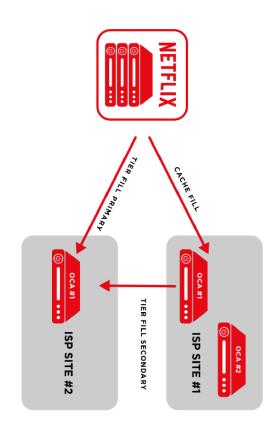
server (data center). See https://openconnect.netflix.com/deploymentguide.pdf.	terminal; and wherein	
terminal" in this example that is served data from the relay server (OCA#1) retrieved from the first	least one target	
server" to fill cached content to each other, from OCA#1 to OCA#2. OCA#2 becomes a "target	first server to at	
peer filling content "cache fill" from the first server to OCA #1. OCA#1 then acts as a "relay	retrieved from said	
According to Netflix' network architecture, OCAs in a cluster and on the same subnet can attempt	serving data	
	relay servers for	
server to at least one OCA terminal.	adapted to act as	
least two of the OCAs are adapted to act as relay servers for serving data retrieved from said first	terminals are	
A "first server" or S3 is identified above in Sec. [1a] as a server within Netflix' data centers, and at	at least two of said	[1d]
Saving this information to a database is common practice within network management.		
should wait for a response before escalation to the next download locations.		
performance of the download location can take on additional traffic or not and how long the OCA		
includes OCAs, and uses a fill escalation policy based upon response times to determine if the		
The CCS server monitors the "fill health" of each potential download locations, or fill sources that		



 $\infty$ 

which then acts as a "relay server" to fill a target terminal (OCA#2) using "tier filling." In another example, Netflix's first server (data center) can cache fill content to a terminal OCA#1,

cluster or subnet) will consider filling from each other rather than via a regular cache fill." considered. OCAs that can see each other's IP address in their BGP feed (but are not in the same "Peer filling is the most efficient method, however as a second-best option tier filling is



See https://openconnect.netflix.com/deploymentguide.pdf.

servers, or "masters" that target terminals can use to gain, or fill, content: "The control plane elects the specified number of OCAs as masters...." OCAs can act as relay In another example, https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states that there are

Title
(content)
availability -
Title (content) availability — Does the fill source have the requested title stored.

- Fill health Can the fill source take on additional fill traffic?
- A calculated route cost Described in the next section.

## Calculating the Least Expensive Fill Source

route possible. the title is passed from one part of our network to another using the most efficient from S3 to all of our OCAs, so we use a tiered approach. The goal is to ensure that It would be inefficient, in terms of both time and cost, to distribute a title directly

Operations team. For example: some configuration parameters for each OCA that are set by the Open Connect To calculate the least expensive fill source, we take into account network state and

- BGP path attributes and physical location (latitude / longitude)
- Fill master (number per fill cluster)
- Fill escalation policies

A fill escalation policy defines:

- 1. How many hops away an OCA can go to download content, and how long it should wait before doing so
- 2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so
- 3. Whether the OCA can go to S3, and how long it should wait before doing so

asset. The fill escalation policies that are applied to masters typically allow them to The control plane elects the specified number of OCAs as masters for a given title

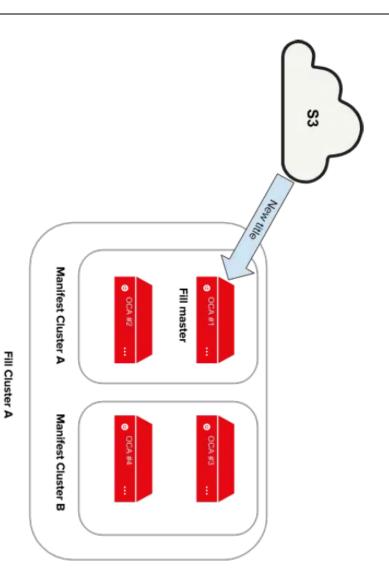
serve this new title have it stored. time zones — until enough of the OCAs in our global network that need to be able to time zone ends, and the fill pattern continues as the fill window moves across other that ensures we don't interrupt any live sessions. needed, they are put into a delete manifest and then deleted after a period of time during the fill window. If there are titles being stored on an OCA that are no longer status, other OCAs can then fill from them, and so on. This process continues As the sun moves west and more members begin streaming, the fill window in this When the second tier of OCAs complete their download, they report back their

## 2. Tier fill: Available OCAs outside the manifest cluster configuration 1. Peer fill: Available OCAs within the same manifest cluster or the same subnet

- Cache fill: Direct download from S3

### Example Scenario

given the option to fill from the fill master. communicate with the control plane to request a fill source for this title, they are control plane that it now has the title stored. The next time the other OCAs After the fill master OCA has completed its S3 download, it reports back to the



the main server is adapted to send transport requests direct to at least one first target	basis of said terminal performance information, an wherein the ma	basis termi perfor inforr where server adapt respor termi netwo which are se as rela	basis termi perfor inforr where server adapt respo termi netwo which are se as rel; a part transf	basis termi perfor inforr where server adapt respor termi netwo which are se as relations times target
All OCA deployments are constantly monitored to ensure reliability and efficiency. Netflix makes use of non-peak bandwidth to download the vast majority of content updates to the OCAs in network during these configurable time windows. OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle.  A "desired manifest" and when needed an "emergency manifest" are posted by Netflix to the CCS server in AWS. These manifests are data structures that indicate what titles need to be transferred to a specific OCA terminal, either during an off-peak hours download window, or urgently if there is an emergent need.		OCAs do not store any information about other OCAs in the network, title popularity, etc. All of this information is aggregated and stored in the AWS control plane. OCAs communicate at regular intervals with the control plane services, requesting (among other things) a manifest file that contains the list of titles they should be storing and serving to members. If there is a delta between the list of titles in the manifest and what they are currently storing, each OCA will send a request,	CAs do not store any information about other OCAs in the network, title opularity, etc. All of this information is aggregated and stored in the AWS control lane. OCAs communicate at regular intervals with the control plane services, equesting (among other things) a manifest file that contains the list of titles they hould be storing and serving to members. If there is a delta between the list of titles in the manifest and what they are currently storing, each OCA will send a request, uring its configured fill window, that includes a list of the new or updated titles	ill Source Manifests  CAs do not store any information about other OCAs in the network, title opularity, etc. All of this information is aggregated and stored in the AWS control lane. OCAs communicate at regular intervals with the control plane services, equesting (among other things) a manifest file that contains the list of titles they hould be storing and serving to members. If there is a delta between the list of titles a the manifest and what they are currently storing, each OCA will send a request, uring its configured fill window, that includes a list of the new or updated titles ontrol Content Service) server in AWS. These manifests are data structures that indicate what les need to be transferred to a specific OCA terminal, either during an off-peak hours download
wherein the main server is further  adapted to monitor  OCAs do not sto		ict	for	rst (ct
of tor	of	ct for	for	for for st
of tor	-	for	for	for contract of the contract o
oct of other	· · · · · · · · · · · · · · · · · · ·			ts:

local OCA terminal hard drive.

 $\triangleright$ Netflix wrote: of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a>, and in which terminals are selected to act as relay servers for a particular data transfers on the basis server/CCS server to a terminal/OCA to download content data in the manifest that needs to be posts a desired manifest or emergency manifest to the CCS server that is intended for each OCA in direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix substantially the same. The claim states that the main server is adapted to send transport requests a terminal (OCA) with instructions to download or "fill" content. The way this is accomplished is The response from the control plane in AWS is a ranked list of potential download locations, a.k.a terminals as a download location for a given title asset. the network. The results are substantially the same - to transmit commands from the main terminal." The functions are substantially the same - to send a request from a main server (CCS) to equivalent the "server is adapted to send transport requests direct to at least one first target fill sources, for each title. The determination of the list takes into consideration several high-level The main server on CCS is further adapted to monitor response times of terminals in the network https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 Thus, the CCS server selects OCA The control plane elects the specified number of OCAs as masters for a given title asset This desired manifest and emergency manifest and the download location data on the CCS fill escalation policy defines: How many hops away an OCA can go to download content, and how long it should wait Fill health — Can the fill source take on additional fill traffic? A calculated route cost — Described in the next section Title (content) availability — Does the fill source have the requested title stored? are

before doing so

OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle. The OCAs work in a network to distribute updates
This desired manifest and emergency manifest along with the downloaded location information files and fill policy for master OCAs and a second (target terminal) OCA to fill from a (first terminal) OCA on the CCS are equivalent the "each such transport request includes details of data to be retrieved, the address of the first server from which the data is to be requested by the first server to be relayed by the first target terminal." The functions are substantially the same - to send a request from a main server (CCS) to a terminal (OCA) with instructions to download or "fill" content from specific master terminal/OCA addresses and includes an address of at least one second terminal/OCA. The way this is accomplished is substantially the same. The claim states that the main server is adapted to send transport requests that include download locations direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix posts a desired for each OCA to read on a regular basis. The results are substantially the same - to transmit commanifest that needs to be retrieved from various addresses.
As stated in [1e], the CCS monitors the "fill health" and performance of download locations, which is based on performance of the OCAs, to determine if that OCA will be selected as a download location or not.
The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to reach farther with less delay in order to grab that content and then share it locally with non-masters.
performance information stored 3. Whether the OCA can go to S3, and how long it should wait before doing so
2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so
1. How many hops away an OCA can go to download content, and how long it should wait before doing so

target terminals request to selected modified transport

from a set of target

transmit the

relay server; and request is to act as modified transport further target addresses of further includes

the recipient of the terminals for which the modified request, wherein in the transport terminals identified

transport request

servers and to main server or received from the transport requests adapted to modify from other relay relay servers are

Page 19 of 151

Connect Overview, p. 5; Fill Patterns, pp. 1-3 among each other and to include further OCAs to which updates and content can be sent. See Open

transport request to selected target terminals that includes addresses of further target terminals requests received from the main server or from other relay servers and transmit the modified Netflix' OCA that are adapted to act as relay servers (see 1d above) are adapted to modify transport

of these master or target terminals in the desired manifest, which is loaded by an OCA terminal in discriminator; 4) the geographically closest appliance. The CCS server includes the URL addresses terminal appliance that receives the route to the client's netblock with the lowest multi-exit the to transfer titles from: 1) the terminal appliance that receives the most-specific route; 2) the its memory or hard drive space in order to select an OCA for downloading titles from terminal appliance that receives the route to the client's netblock with the shortest AS path; 3) the CCS server. The CCS server uses Appliance Section Criteria to select OCA terminals as targets in The CCS server will to order OCA terminals to peer or tier fill from using OCAs selected by the

locations of master OCAs for each individual title needed by an OCA to fill its delta: on the delta. The CCS responds, as stated in [1f] with a list of URLs that are downloadable from the actual manifest are termed the "delta" or difference between actual and desired manifest. "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing The OCA terminal will then query the CCS terminal for a list of download locations for each title After receiving the desired manifest, an OCA terminal will compare the manifest with its own

things) a manifest file that contains the list of titles they should be storing and serving to members https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 potential download locations, a.k.a. fill sources, for each title." (Emphasis added.) See updated titles that it needs. The response from the control plane in AWS is a ranked list of If there is a **delta** between the list of titles in the manifest and what they are currently storing, each "OCAs communicate at regular intervals with the control plane services, requesting (among other OCA will send a request, during its configured fill window, that includes a list of the new or

and then requesting a delta-listed title from the list of a master OCAs, is equivalent to a modified The action of an OCA requesting download locations (master OCAs) for its delta list from the CCS

desired manifest) to request a title or titles from further target terminals, or master OCAs. transport request. The OCA is using a modified list of titles (delta or missing titles list from the

substantially similar: further relay terminal addresses are sent to the OCAs, a modification of the of URL locations to download the modified list titles of its delta list, after which the delta list is another OCA, an "actual manifest" is sent to the CCS server which responds to the OCA with a list of what it is supposed to download. Instead of the terminal, or OCA, transmitting the delta list to substantially the same. A delta is a modification of the manifest list, or in other words a subset list with URLs of those relay servers or URLs of master OCAs. The way the actions are performed are relay servers or OCAs a modification of the original transport request, or desired manifest, along transport request to selected target terminals from a set of target terminals identified in the transport request, wherein the modified transport request further includes addresses of further target terminals requests received from the main server or from other relay servers and to transmit the modified then sent to another OCA via a series of modified transfer requests. The result of these actions are This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport The function of the two actions are substantially the same, which is to transmit to other selected for which the recipient of the modified transport request is to act as relay server."

clustered or if they are in the same subnet, will attempt to peer or tier fill from each other. https://openconnect.zendesk.com/hc/en-us/articles/360035618071-Fill-patterns Alternatively, and equivalently, Netflix documentation discloses that OCA terminals, if they are

original transport request (desired manifest) is transmitted to selected relay terminals (single or

multiple delta titles to selected OCAs) are sent to addresses (URLs) of further selected relay servers

(OCAs) in the form of a request for one or more titles from one or more master OCAs.

save this information. In general, appliances determine where to receive fill using selection criteria geographically closest terminal. The OCA terminal includes the URL addresses of these terminals that receives the route to the client's netblock with the lowest multi-exit discriminator; 4) the titles from: 1) the terminal appliance that receives the most-specific route; 2) the appliance terminal that is used by Netflix client devices. The OCA terminals then use a similar Appliance Section that receives the route to the client's netblock with the shortest AS path; 3) the appliance terminal Criteria as the CCS server uses to select OCA terminals as targets in the subnet or cluster to transfer OCA terminals in a subnet or cluster broadcast their IP and physical locations to one another and

		[1h]				
said target	target terminals are divided into a series of packets for transmission to	wherein data to be retrieved by said				
	The Open Connect network can make parallelized cache fill transfers. As such, it must be transferring "packets" of the file. There is evidence that each file is downloaded for cache fill in packet sizes up to 16 kilobytes.	Netflix' OCAs are adapted to communicate with the main server which is hosted by AWS. The OCA's "Report their status to the Open Connect control panel services in Amazon Web Services":	The function of the two actions are substantially the same, which is to transmit to other selected relay servers or OCAs a modified list of the original transport request or manifest. The way the actions are performed are substantially the same. A modified desired manifest list is a subset list called the "delta." A relay server (OCA) creates a modified transport request (delta list) of its missing titles and transports this list to another relay server (OCA sends requests to other selected OCAs for the titles on the delta list in a series of requests) and includes addresses of the other relay servers (URLs of the selected OCA terminals in the request as an address to the selected OCAs). The result of these actions are substantially similar: a modification of the original transport request (desired manifest) is transmitted to selected relay terminals (selected OCAs) with addresses (URLs) of the selected relay servers (OCAs).	This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport requests received from the main server or from other relay servers and to transmit the modified transport request to selected target terminals from a set of target terminals identified in the transport request, wherein the modified transport request further includes addresses of further target terminals for which the recipient of the modified transport request is to act as relay server."	After receiving the desired manifest, an OCA terminal will compare the manifest with its own "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing from the actual manifest are termed the "delta" or difference between actual and desired manifest. After selecting an OCA master using the selection criteria, the OCA terminal will transmit and request the delta list items to the selected OCA(s) in the subnet or cluster in the form of download requests for each title using the URL of the target OCA.	in its memory or hard drive space in order to select an OCA for downloading titles from. See <i>Fill Patterns</i> , pp. 1-4.

					1
			acknowledge receipt of the last packet of a series routed thereto.	directly with said main server to	terminals and each of said terminals is adapted to communicate
<ul> <li>Traffic from OCA: Allow all destination addresses and ports.</li> <li>Traffic to OCA: Allow TCP 22, 53, 80, 179, 443, UDP 53 and 123 (source and destination), ICMP types 0, 3, 8, 11, and all ICMPv6 from any public IP/port. Allow all return traffic from any appliance-initiated connection (TCP established).</li> <li>https://openconnect.zendesk.com/hc/en-us/articles/360035533071#routi</li> </ul>	Comer, Douglas E. (2006). Internetworking with TCP/IP: Principles, Protocols, and Architecture. Vol. 1 (5th ed.). Prentice Hall. ISBN 978-0-13-187671-2  Netflix requires network traffic to OCA be in TCP protocol:	TCP is a reliable byte stream delivery service which guarantees that all bytes received will be identical and in the same order as those sent. Since packet transfer by many networks is not reliable, TCP achieves this using a technique known as positive acknowledgement with retransmission. This requires the receiver to respond with an acknowledgement message as it receives the data. The sender keeps a record of each packet it sends and maintains a timer from when the packet was sent. The sender re-transmits a packet if the timer expires before receiving the acknowledgement. The timer is needed in case a packet gets lost or corrupted.	TCP data communication protocol, which most of the Internet included Netflix uses, requires an acknowledgement of packets when content is transmitted:	There is also evidence that the cache fill transfer uses TCP/IP protocols for packetized data transfer.	Additionally, the last packet of a download for a cache fill is associated with a notice to the CCS that the download has been completed. The system can also use parallel processes to perform cache transfer using TCP/IP protocols. This is only possible if the CCS is working on a different section of the file at the same time, which is packetized file transfer.

### 37

## Glossary of terms

Compound Annual Growth Rate

A temporary local copy of information that in a local server of files to be delivered to consumers, stored originated elsewhere. Thus for CDNs, a copy

## **CDN (Content Delivery Network)**

content over the internet enable the efficient and reliable distribution of A distributed system of servers, designed to

### CP (content provider)

content (rather than - say - e-commerce) An internet business whose focus is delivering to consumers. Netflix, CNN and YouTube

## across multiple servers

transmits content

One step in a packet of data's journey

Fixed broadband

to consumers (consumers or businesses). May ISP (Internet Service Provider) A company providing internet connectivity

TCP/IP

### IXP (Internet Exchange Point)

provide fixed broadband, mobile data or both

data transmission that underpin the internet protocol. The two foundational standards for

Transmission control protocol and internet

to exchange traffic, avoiding the need A location where many networks meet for multiple bilateral connections

sent and reaching its destination The lag between a packet of data being

### OTT (Over The Top)

Describes services delivered over another

broadband networks) (since they are not provided by telcos operating YouTube, Facebook and Netflix are examples network without being integrated with it.

### Packet loss

from the source server data in question will then be requested again This is known as packet loss. Typically the handle, it discards a certain amount of data When a router is sent more data than it can

### Server

of data and sends them onwards down the A computer that stores and appropriate link A switch on the internet, that receives packets

		"A cooperative approach to content delivery," Netflix (2021), 37.
		Where does Netflix use TCP?
		Netflix uses TCP for internet streaming to send packets of data for video. Additionally, Netflix specifically looks at the number of TCP connections to determine internet speeds in accordance with testing of OCAs.
		In another example, "After <b>the fill master OCA</b> has completed its S3 download, <b>it reports back to the control plane</b> that it now has the title stored." See https://netflixtechblog.com/netflix-and-fill-c43a32b490c0
[3a]	The network as claimed in claim 1,	See above.
[3b]	wherein terminals acting as relay servers are adapted to select further downstream target terminals to act as further relay servers on the basis of their relative performances of the further target terminals indicated in said transport request.	OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle. The OCAs work in a network to distribute updates among each other and to include further OCAs to which updates and content can be sent. See Open Connect Overview, p. 5.  Alternatively, and equivalently, Netflix documentation discloses that OCA terminals, if they are clustered or if they are in the same subnet, will attempt to peer or tier fill from each other. https://openconnect.zendesk.com/hc/en-us/articles/360035618071-Fill-patterns  OCA terminals in a subnet or cluster broadcast their IP and physical locations to one another and save this information. In general, appliances determine where to receive fill using selection criteria that is used by Netflix client devices. The OCA terminals then use a similar Appliance Section Criteria as the CCS server uses to select OCA terminals as targets in the subnet or cluster to transfer titles from: 1) the terminal appliance that receives the most-specific route; 2) the appliance terminal that receives the route to the client's netblock with the shortest AS path; 3) the appliance terminal that receives the route to the client's netblock with the lowest multi-exit discriminator; 4) the

requests for each title using the URL of the target OCA. request the delta list items to the selected OCA(s) in the subnet or cluster in the form of download After selecting an OCA master using the selection criteria, the OCA terminal will transmit and "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing in its memory or hard drive space in order to select an OCA for downloading titles from. See Fill geographically closest terminal. The OCA terminal includes the URL addresses of these terminals Patterns, pp. 1-4. from the actual manifest are termed the "delta" or difference between actual and desired manifest. After receiving the desired manifest, an OCA terminal will compare the manifest with its own

• Fill health — Can the fill source take on additional fill traffic?

A calculated route cost — Described in the next section.

## Calculating the Least Expensive Fill Source

It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to *all* of our OCAs, so we use a tiered approach. The goal is to ensure that the title is passed from one part of our network to another using the most efficient route possible.

To calculate the least expensive fill source, we take into account network state and some configuration parameters for each OCA that are set by the Open Connect Operations team. For example:

- BGP path attributes and physical location (latitude / longitude)
- Fill master (number per fill cluster)
- Fill escalation policies

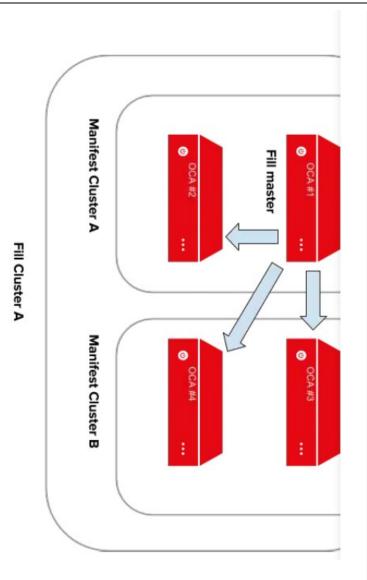
A fill escalation policy defines:

- How many hops away an OCA can go to download content, and how long it should wait before doing so
- 2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so
- 3. Whether the OCA can go to S3, and how long it should wait before doing so

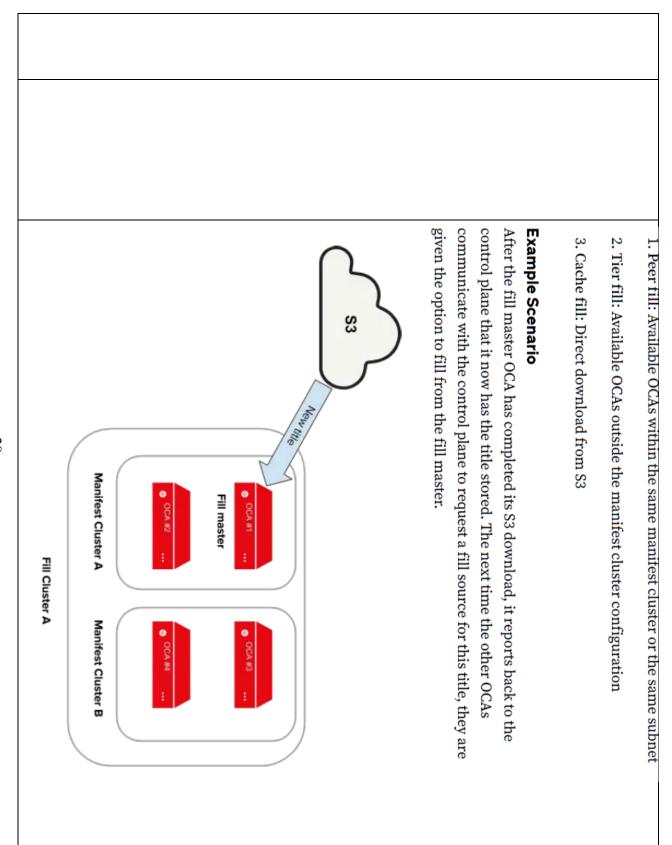
The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to

## Fill Source Manifests

plane. OCAs communicate at regular intervals with the control plane services, during its configured fill window, that includes a list of the new or updated titles in the manifest and what they are currently storing, each OCA will send a request, should be storing and serving to members. If there is a delta between the list of titles requesting (among other things) a manifest file that contains the list of titles they popularity, etc. All of this information is aggregated and stored in the AWS control OCAs do not store any information about other OCAs in the network, title



serve this new title have it stored. time zones — until enough of the OCAs in our global network that need to be able to time zone ends, and the fill pattern continues as the fill window moves across other that ensures we don't interrupt any live sessions. needed, they are put into a delete manifest and then deleted after a period of time during the fill window. If there are titles being stored on an OCA that are no longer status, other OCAs can then fill from them, and so on. This process continues As the sun moves west and more members begin streaming, the fill window in this When the second tier of OCAs complete their download, they report back their



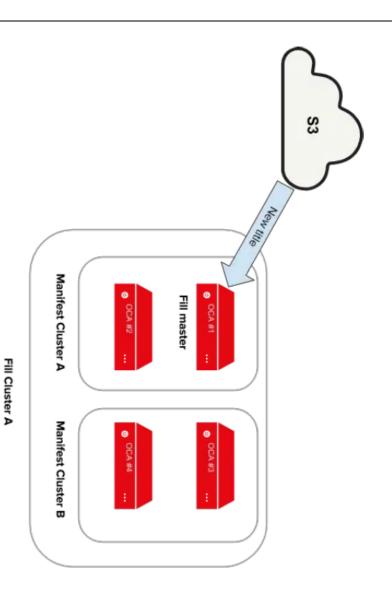
		The control panel on the main server observes performance data to determine which "second tier" OCAs to complete their downloads with from a master OCA, to "report back their (second tier OCAs) status" to the main server, "and so on" to send transport requests to each new "master" OCA and new "second tier" OCA's, "and so on."
[4a]	The network as claimed in claim 1,	See above.
[4b]	wherein the first server is a terminal adapted to act as relay server.	In the Netflix architecture diagram, a terminal can act as the first server, where the other OCAs can have the option of downloading from the "fill master OCA."

# 1. Peer fill: Available OCAs within the same manifest cluster or the same subnet

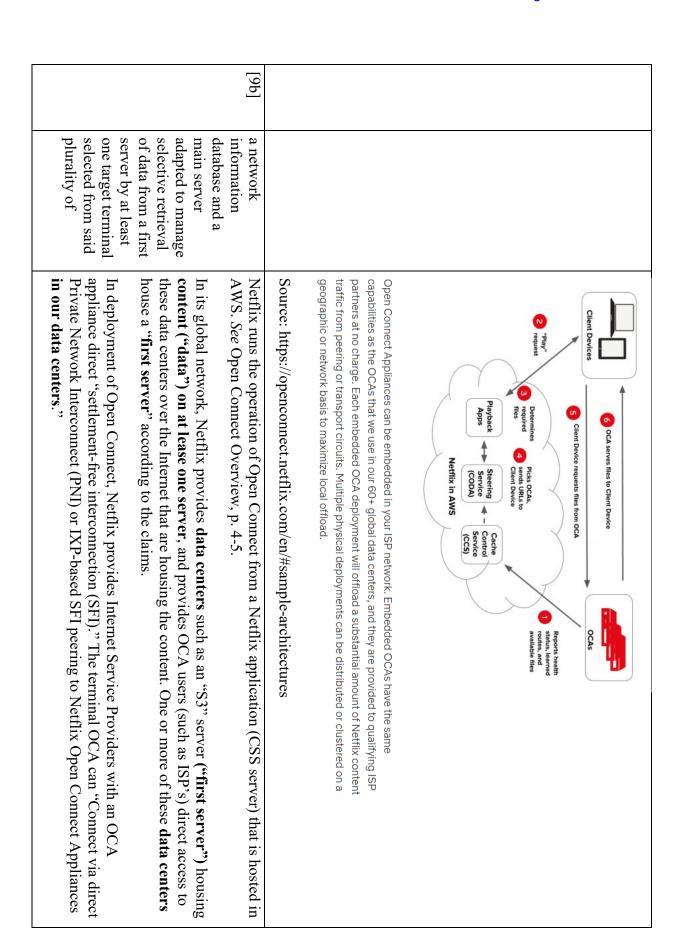
2. Tier fill: Available OCAs outside the manifest cluster configuration

- Cache fill: Direct download from S3

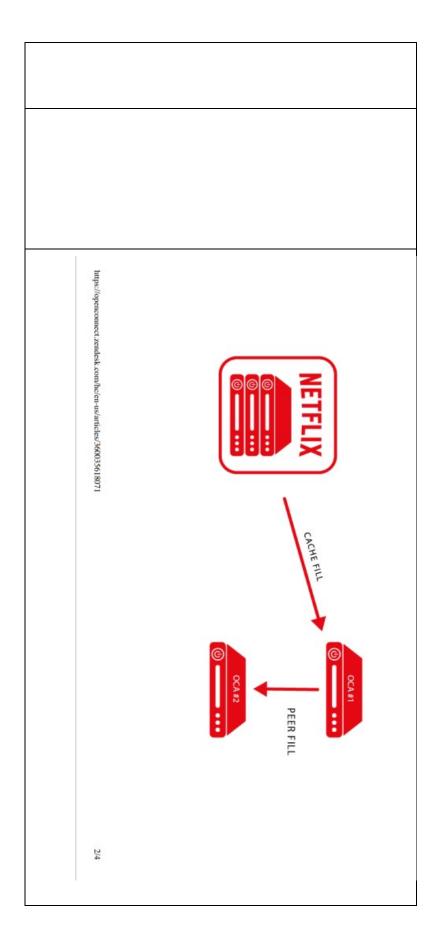
given the option to fill from the fill master. communicate with the control plane to request a fill source for this title, they are control plane that it now has the title stored. The next time the other OCAs After the fill master OCA has completed its S3 download, it reports back to the Example Scenario

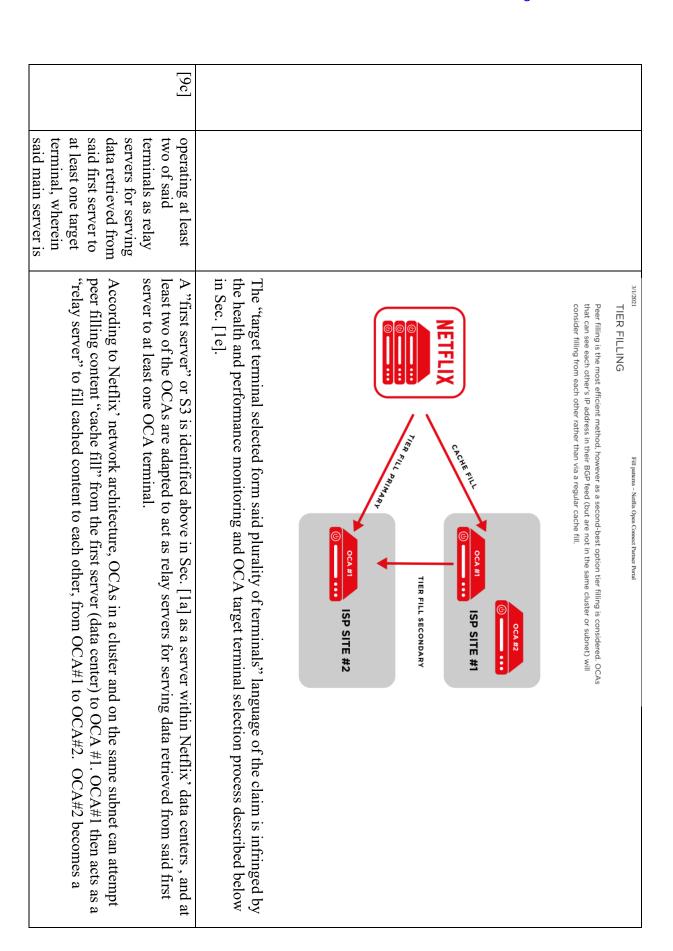


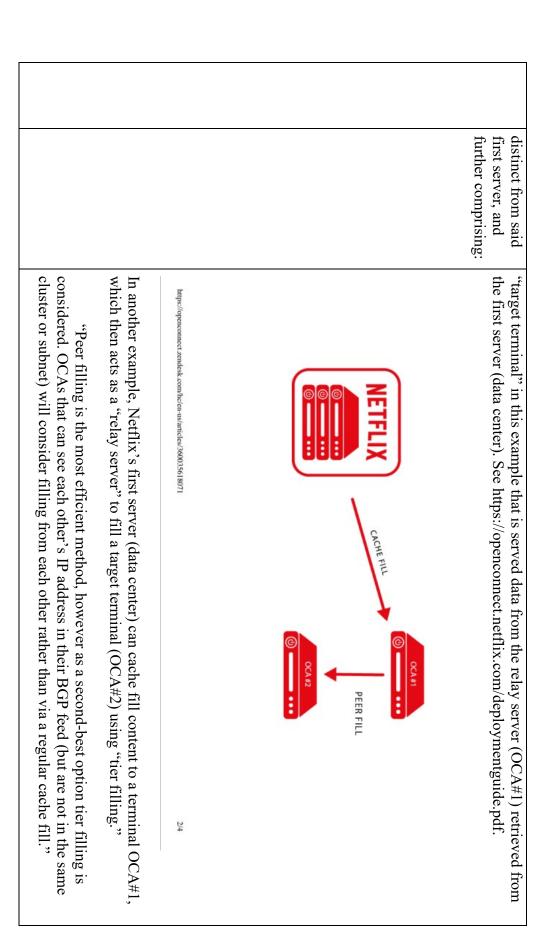
Netflix uses a method to deliver Netflix TV shows and movies to members using world-wide system called <b>Open Connect</b> .  The building blocks of Open Connect are our suite of purpose-built <b>server appliances, called Open Connect Appliances (OCAs)</b> . See Open Connect Overview, p. 2. These are deployed directly inside ISP networks. Netflix provides the server hardware. The OCAs report to a Open Connect control plane to control fill behavior (adding new files to OCAs nightly) and to compute and/or store data. See id. p. 3-4. Accordingly, OCAs include both an input mechanism and display mechanism.	A method of operating a data communication network, the data communication network comprising: a plurality of terminals	[9a]	
ICC Profile — The information needed to define and convert between color profiles (color primaries, tone reproduction curve, absolute luminance) can be encoded in an ICC Profile, which is an ISO-standardized binary file format that can be embedded in many image file formats. It's also natively supported by industry standard tools like Adobe Photoshop, and by operating systems like Windows and macOS.			
Solving for the HDR Image Format  After exploring all of the requirements above we ultimately came to the realization that many common image file formats that have existed for years can already technically support the characterization of HDR pixels, but there are trade-offs for each one. The key realization was that we could add an ICC Profile to each image to signal the HDR color profile of that image.			
Netflix states, on its public technical blog, that binary format data can be used in image file formats sent through the network. <i>See</i> <a href="https://netflixtechblog.com/enhancing-the-netflix-ui-experience-with-hdr-1e7506ad3e8">https://netflixtechblog.com/enhancing-the-netflix-ui-experience-with-hdr-1e7506ad3e8</a>	wherein data is transmitted in binary format.	[8b]	



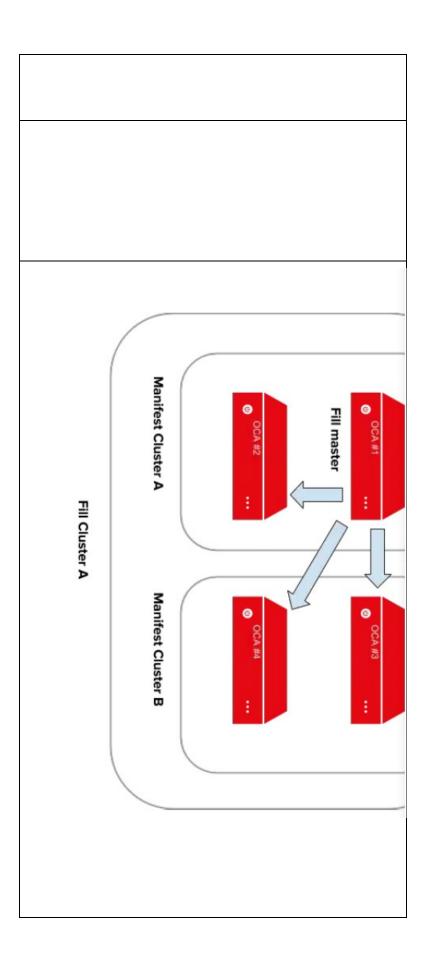
				terminals; comprising
OCAs in a cluster and on the same subnet can attempt peer filling from each other. There is also Tier filling where if in different ISP sites. <a href="deploymentguide.pdf">deploymentguide.pdf</a> (netflix.com)	The following diagram also illustrates access from a target terminal (OCA $\#1$ , OCA $\#2$ ) to a Netflix first server in "our data centers."	Welcome to Open Connect, p. 3. Dkt 39 at p. 48.	ISPs who do not currently participate in pubic peering might want to consider that a single IX port can support multiple peering sessions, providing <b>direct access to various content</b> , cloud, and network providers.	"Netflix has the ability to interconnect at a number of global data center facilities and public Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any network at IXP locations where we are mutually present and we consider private interconnection as appropriate."







## should be storing and serving to members. If there is a delta between the list of titles during its configured fill window, that includes a list of the new or updated titles in the manifest and what they are currently storing, each OCA will send a request, requesting (among other things) a manifest file that contains the list of titles they plane. OCAs communicate at regular intervals with the control plane services, popularity, etc. All of this information is aggregated and stored in the AWS control OCAs do not store any information about other OCAs in the network, title Fill Source Manifests

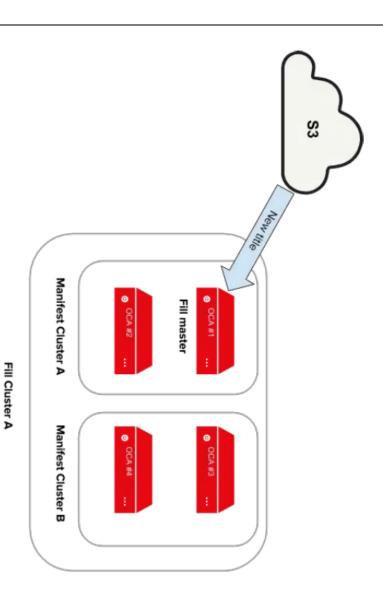


serve this new title have it stored. time zones — until enough of the OCAs in our global network that need to be able to time zone ends, and the fill pattern continues as the fill window moves across other that ensures we don't interrupt any live sessions. needed, they are put into a delete manifest and then deleted after a period of time during the fill window. If there are titles being stored on an OCA that are no longer status, other OCAs can then fill from them, and so on. This process continues As the sun moves west and more members begin streaming, the fill window in this When the second tier of OCAs complete their download, they report back their

- 1. Peer fill: Available OCAs within the same manifest cluster or the same subnet
- 2. Tier fill: Available OCAs outside the manifest cluster configuration
- 3. Cache fill: Direct download from S3

## Example Scenario

given the option to fill from the fill master. communicate with the control plane to request a fill source for this title, they are control plane that it now has the title stored. The next time the other OCAs After the fill master OCA has completed its S3 download, it reports back to the



	[9d]
	sending transport requests from the main server to at least one first target terminal based on terminal performation stored in the network information database; and operating the first target terminal to act as relay server;
This desired manifest and emergency manifest and the download location data on the CCS are equivalent the "server is adapted to send transport requests direct to at least one first target terminal." The functions are substantially the same - to send a request from a main server (CCS) to a terminal (OCA) with instructions to download or "fill" content. The way this is accomplished is substantially the same. The claim states that the main server is adapted to send transport requests direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix posts a desired manifest or emergency manifest to the CCS server that is intended for each OCA in the network. The results are substantially the same - to transmit commands from the main server/CCS server to a terminal/OCA to download content data in the manifest that needs to be retrieved.	All OCA deployments are constantly monitored to ensure reliability and efficiency. Netflix makes use of non-peak bandwidth to download the vast majority of content updates to the OCAs in network during these configurable time windows. OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle.  A "desired manifest" and when needed an "emergency manifest" are posted by Netflix to the CCS server in AWS. These manifests are data structures that indicate what titles need to be transferred to a specific OCA terminal, either during an off-peak hours download window, or urgently if there is an emergent need.  A "desired manifest" and when needed an "emergency manifest" are posted by Netflix to the CCS (Control Content Service) server in AWS. These manifests are data structures that indicate what titles need to be transferred to a specific OCA terminal, either during an off-peak hours download window, or urgently if there is an emergent need. Equivalently, these manifests are fetched by regularly from the CCS by each OCA terminal, which traverses the desired manifest or emergency manifest. The CCS server acts as an email inbox for manifests sent directly to the OCAs, where each OCA terminal hard drive.

terminals as a download location for a given title asset https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 Thus, the CCS server selects OCA The control plane elects the specified number of OCAs as masters for a given title asset

of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a> and in which terminals are selected to act as relay servers for a particular data transfers on the basis Netflix wrote: The main server on CCS is further adapted to monitor response times of terminals in the network

fill sources, for each title. The determination of the list takes into consideration several high-level The response from the control plane in AWS is a ranked list of potential download locations, a.k.a

Title (content) availability — Does the fill source have the requested title stored?

# Fill health — Can the fill source take on additional fill traffic?

A calculated route cost — Described in the next section

A fill escalation policy defines:

- How many hops away an OCA can go to download content, and how long it should wait before doing so
- above), and how long it should wait before doing so Whether the OCA can go to the entire Open Connect network (beyond the hops defined

 $\infty$ 

Whether the OCA can go to S3, and how long it should wait before doing so

## (Emphasis added.)

assembled by the CCS server and download locations are assembled based on OCA performance should wait for a response before escalation to the next download locations. Thus, the manifest is includes OCAs, and uses a fill escalation policy based upon response times to determine if the performance of the download location can take on additional traffic or not and how long the OCA The CCS server monitors the "fill health" of each potential download locations, or fill sources that

of at least one second target terminal to which the data retrieved	וווכ ממומ זכוווכעכם	from the first server is to be all	first l and of a	р В #	e first s to be by the first rminal and ation of a ance of a larget l based on inal	e first s to be by the first rminal and ation of a ance of a target l based on inal ance
https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one second OCA can be selected based on their relative performance:	https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one second OCA can be selected based on their relative performance:  The CCS server information location list provides the address of a first server, called the "S3" server, for a download location and provides other download locations of a second and additional master OCAs:	https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one second OCA can be selected based on their relative performance:  The CCS server information location list provides the address of a first server, called the "S3" server, for a download location and provides other download locations of a second and additional master OCAs:  It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to all of our OCAs, so we use a tiered approach A fill escalation policy defines:	https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one second OCA can be selected based on their relative performance:  The CCS server information location list provides the address of a first server, called the "S3" server, for a download location and provides other download locations of a second and additional master OCAs:  It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 t all of our OCAs, so we use a tiered approach A fill escalation policy defines:  1. How many hops away an OCA can go to download content, and how long it should wait before doing so	https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one second OCA can be selected based on their relative performance:  The CCS server information location list provides the address of a first server, called the "S3" server, for a download location and provides other download locations of a second and additional master OCAs:  It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to all of our OCAs, so we use a tiered approach A fill escalation policy defines:  1. How many hops away an OCA can go to download content, and how long it should wait before doing so  2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so	//netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one id OCA can be selected based on their relative performance:  CCS server information location list provides the address of a first server, called the "S3" r, for a download location and provides other download locations of a second and additional rr OCAs:  It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 t our OCAs, so we use a tiered approach A fill escalation policy defines:  1. How many hops away an OCA can go to download content, and how long it should before doing so  2. Whether the OCA can go to the entire Open Connect network (beyond the hops defines), and how long it should wait before doing so  3. Whether the OCA can go to S3, and how long it should wait before doing so	//netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one d OCA can be selected based on their relative performance:  CCS server information location list provides the address of a first server, called the "S3" r, for a download location and provides other download locations of a second and additional rr OCAs:  It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 tour OCAs, so we use a tiered approach A fill escalation policy defines:  1. How many hops away an OCA can go to download content, and how long it should before doing so  2. Whether the OCA can go to the entire Open Connect network (beyond the hops defines), and how long it should wait before doing so  3. Whether the OCA can go to S3, and how long it should wait before doing so

	Case	4:22-C	V-0149	9U-JS	Filed 06/16/23 Page 49	9 OT 151
				[9i]		
transmit the	from other relay servers and to	received from the	relay servers are adapted to modify	operating terminals		in the network information database;
	transport request to selected target terminals that includes addresses of further target terminals.	Netflix' OCA that are adapted to act as relay servers (see 1d above) are adapted to modify transport requests received from the main server or from other relay servers and transmit the modified	among each other and to include further OCAs to which updates and content can be sent. See Open Connect Overview, p. 5; Fill Patterns, pp. 1-3.	OCAs can also download updates from each other – minimizing significant usage of internet "hackbone" canacity during the undates excle. The OCAs work in a network to distribute undates	is based on performance of the OCAs, to determine if that OCA will be selected as a download location or not.  This desired manifest and emergency manifest along with the downloaded location information files and fill policy for master OCAs and a second (target terminal) OCA to fill from a (first terminal) OCA on the CCS are equivalent the "each such transport request includes details of data to be retrieved, the addresses of the first server from which the data is to be requested by the first target terminal, the addresses of at least one second target terminal to which the data from the first server to be relayed by the first target terminal." The functions are substantially the same - to send a request from a main server (CCS) to a terminal (OCA) with instructions to download or "fill" content from specific master terminal/OCA addresses and includes an address of at least one second terminal/OCA. The way this is accomplished is substantially the same. The claim states that the main server is adapted to send transport requests that include download locations direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix posts a desired manifest or emergency manifest to the CCS server along with download locations that are intended for each OCA to read on a regular basis. The results are substantially the same - to transmit commands from the main server/CCS server to a terminal/OCA to download content data in the manifest that needs to be retrieved from various addresses.	The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to reach farther with less delay in order to grab that content and then share it locally with non-masters.  As stated in [1e] the CCS monitors the "fill health" and performance of download locations which

request is to act as modified transport

relay server; and

terminals for which

the recipient of the

further target addresses of further includes the modified

transport request

in the transport

request, wherein terminals identified target terminals from a set of target request to selected modified transport of these master or target terminals in the desired manifest, which is loaded by an OCA terminal in discriminator; 4) the geographically closest appliance. The CCS server includes the URL addresses terminal appliance that receives the route to the client's netblock with the lowest multi-exit terminal appliance that receives the route to the client's netblock with the shortest AS path; 3) the CCS server. The CCS server uses Appliance Section Criteria to select OCA terminals as targets in The CCS server will to order OCA terminals to peer or tier fill from using OCAs selected by the the to transfer titles from: 1) the terminal appliance that receives the most-specific route; 2) the its memory or hard drive space in order to select an OCA for downloading titles from.

on the delta. The CCS responds, as stated in [1f] with a list of URLs that are downloadable The OCA terminal will then query the CCS terminal for a list of download locations for each title "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing locations of master OCAs for each individual title needed by an OCA to fill its delta: from the actual manifest are termed the "delta" or difference between actual and desired manifest. After receiving the desired manifest, an OCA terminal will compare the manifest with its own

things) a manifest file that contains the list of titles they should be storing and serving to members https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 potential download locations, a.k.a. fill sources, for each title." (Emphasis added.) See updated titles that it needs. The response from the control plane in AWS is a ranked list of If there is a **delta** between the list of titles in the manifest and what they are currently storing, each "OCAs communicate at regular intervals with the control plane services, requesting (among other OCA will send a request, during its configured fill window, that includes a list of the new or

and then requesting a delta-listed title from the list of a master OCAs, is equivalent to a modified desired manifest) to request a title or titles from further target terminals, or master OCAs. transport request. The OCA is using a modified list of titles (delta or missing titles list from the The action of an OCA requesting download locations (master OCAs) for its delta list from the CCS.

requests received from the main server or from other relay servers and to transmit the modified transport request to selected target terminals from a set of target terminals identified in the transport This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport

request, wherein the modified transport request further includes addresses of further target terminals for which the recipient of the modified transport request is to act as relay server."

original transport request (desired manifest) is transmitted to selected relay terminals (single or substantially similar: further relay terminal addresses are sent to the OCAs, a modification of the of URL locations to download the modified list titles of its delta list, after which the delta list is another OCA, an "actual manifest" is sent to the CCS server which responds to the OCA with a list of what it is supposed to download. Instead of the terminal, or OCA, transmitting the delta list to substantially the same. A delta is a modification of the manifest list, or in other words a subset list with URLs of those relay servers or URLs of master OCAs. The way the actions are performed are relay servers or OCAs a modification of the original transport request, or desired manifest, along multiple delta titles to selected OCAs) are sent to addresses (URLs) of further selected relay servers then sent to another OCA via a series of modified transfer requests. The result of these actions are (OCAs) in the form of a request for one or more titles from one or more master OCAs The function of the two actions are substantially the same, which is to transmit to other selected

clustered or if they are in the same subnet, will attempt to peer or tier fill from each other https://openconnect.zendesk.com/hc/en-us/articles/360035618071-Fill-patterns Alternatively and equivalently, Netflix documentation discloses that OCA terminals, if they are

geographically closest terminal. The OCA terminal includes the URL addresses of these terminals titles from: 1) the terminal appliance that receives the most-specific route; 2) the appliance terminal save this information. In general, appliances determine where to receive fill using selection criteria Patterns, pp. 1-4. in its memory or hard drive space in order to select an OCA for downloading titles from. See Fill that receives the route to the client's netblock with the lowest multi-exit discriminator; 4) the that receives the route to the client's netblock with the shortest AS path; 3) the appliance terminal Criteria as the CCS server uses to select OCA terminals as targets in the subnet or cluster to transfer that is used by Netflix client devices. The OCA terminals then use a similar Appliance Section OCA terminals in a subnet or cluster broadcast their IP and physical locations to one another and

"actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing from the actual manifest are termed the "delta" or difference between actual and desired manifest After receiving the desired manifest, an OCA terminal will compare the manifest with its own

[9j]		
wherein dividing data to be retrieved by said target terminals into a series of packets for transmission to said target terminals and wherein each of said terminals communicates directly with said main server to		
Netflix' OCAs are adapted to communicate with the main server which is hosted by AWS. The OCA's "Report their status to the Open Connect control panel services in Amazon Web Services": The Open Connect network can make parallelized cache fill transfers. As such, it must be transferring "packets" of the file. There is evidence that each file is downloaded for cache fill in parallel in packet sizes up to 16 kilobytes.  Additionally, the last packet of a download for a cache fill is associated with a notice to the CCS that the download has been completed. The system can also use parallel processes to perform cache transfer using TCP/IP protocols. This is only possible if the CCS is working on a different section of the file at the same time, which is packetized file transfer.  There is also evidence that the cache fill transfer uses TCP/IP protocols for packetized data transfer.	requests for each title using the URL of the target OCA.  This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport requests received from the main server or from other relay servers and to transmit the modified transport request to selected target terminals from a set of target terminals identified in the transport request, wherein the modified transport request further includes addresses of further target terminals for which the recipient of the modified transport request is to act as relay server."  The function of the two actions are substantially the same, which is to transmit to other selected relay servers or OCAs a modified list of the original transport request or manifest. The way the actions are performed are substantially the same. A modified desired manifest list is a subset list called the "delta." A relay server (OCA) creates a modified transport request (delta list) of its missing titles and transports this list to another relay server (OCA sends requests to other selected OCAs for the titles on the delta list in a series of requests) and includes addresses of the other relay servers (URLs of these actions are substantially similar: a modification of the original transport request (desired manifest) is transmitted to selected relay terminals (selected OCAs) with addresses (URLs) of the selected relay servers (OCAs).	After selecting an OCA master using the selection criteria, the OCA terminal will transmit and request the delta list items to the selected OCA(s) in the subnet or cluster in the form of download

			acknowledge receipt of the last packet of a series routed thereto.
https://openconnect.zendesk.com/hc/en-us/articles/360035533071#routi	<ul> <li>Traffic from OCA: Allow all destination addresses and ports.</li> <li>Traffic to OCA: Allow TCP 22, 53, 80, 179, 443, UDP 53 and 123 (source and destination), ICMP types 0, 3, 8, 11, and all ICMPv6 from any public IP/port. Allow all return traffic from any appliance-initiated connection (TCP established).</li> </ul>	Netflix requires network traffic to OCA be in TCP protocol:	TCP data communication protocol, which most of the Internet included Netflix uses, requires an acknowledgement of packets when content is transmitted:  TCP is a reliable byte stream delivery service which guarantees that all bytes received will be identical and in the same order as those sent. Since packet transfer by many networks is not reliable TCP achieves this using a technique known as positive acknowledgement with retransmission. This requires the receiver to respond with an acknowledgement message as it receives the data. The sender keeps a record of each packet it sends and maintains a timer from when the packet was sent. The sender re-transmits a packet if the timer expires before receiving the acknowledgement. The timer is needed in case a packet gets lost or corrupted.[  Comer, Douglas E. (2006). Internetworking with TCP/IP: Principles, Protocols, and Architecture. Vol. 1 (5th ed.). Prentice Hall. ISBN 978-0-13-187671-2

#### 37

## Glossary of terms

Compound Annual Growth Rate

A temporary local copy of information that in a local server of files to be delivered to consumers, stored originated elsewhere. Thus for CDNs, a copy

## **CDN (Content Delivery Network)**

CP (content provider) content over the internet enable the efficient and reliable distribution of A distributed system of servers, designed to

content (rather than - say - e-commerce) An internet business whose focus is delivering to consumers. Netflix, CNN and YouTube

transmits content

One step in a packet of data's journey

Fixed broadband

across multiple servers

ISP (Internet Service Provider) A company providing internet connectivity

## IXP (Internet Exchange Point)

provide fixed broadband, mobile data or both to consumers (consumers or businesses). May

data transmission that underpin the internet protocol. The two foundational standards for

Transmission control protocol and internet

for multiple bilateral connections to exchange traffic, avoiding the need A location where many networks meet

sent and reaching its destination The lag between a packet of data being

## OTT (Over The Top)

broadband networks) (since they are not provided by telcos operating YouTube, Facebook and Netflix are examples network without being integrated with it. Describes services delivered over another

## Packet loss

from the source server data in question will then be requested again This is known as packet loss. Typically the handle, it discards a certain amount of data When a router is sent more data than it can

### Server

of data and sends them onwards down the A computer that stores and appropriate link

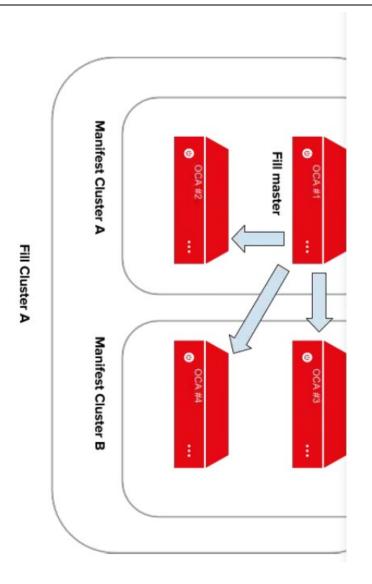
A switch on the internet, that receives packets

#### 53

	the further target terminals indicated in said transport request.	
The document https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states that there are "The control plane elects the specified number of OCAs as masters" OCAs can act as relay servers, or "masters" that target terminals can use to gain, or fill, content and these are based on relative performance of terminals:	terminals to act as further relay servers on the basis of the relative performances of	
OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle. The OCAs work in a network to distribute updates among each other and to include further OCAs to which updates and content can be sent. <i>See</i> Open Connect Overview, p. 5.	including operating terminals acting as relay servers to select further	[116]
See above.	The method as claimed in claim 9,	[11a]
In another example, "After <b>the fill master OCA</b> has completed its S3 download, <b>it reports back to the control plane</b> that it now has the title stored." See https://netflixtechblog.com/netflix-and-fill-c43a32b490c0		
Netflix uses TCP for internet streaming to send packets of data for video. Additionally, Netflix specifically looks at the number of TCP connections to determine internet speeds in accordance with testing of OCAs.		
Where does Netflix use TCP?		
"A cooperative approach to content delivery," Netflix (2021), 37.		

## Fill Source Manifests

should be storing and serving to members. If there is a delta between the list of titles plane. OCAs communicate at regular intervals with the control plane services, during its configured fill window, that includes a list of the new or updated titles in the manifest and what they are currently storing, each OCA will send a request, requesting (among other things) a manifest file that contains the list of titles they popularity, etc. All of this information is aggregated and stored in the AWS control OCAs do not store any information about other OCAs in the network, title



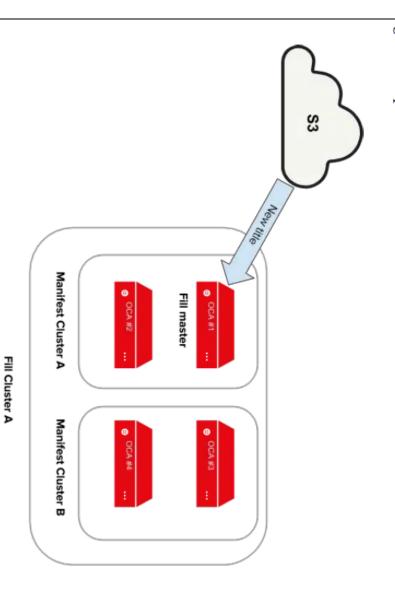
serve this new title have it stored. time zones — until enough of the OCAs in our global network that need to be able to time zone ends, and the fill pattern continues as the fill window moves across other that ensures we don't interrupt any live sessions. needed, they are put into a delete manifest and then deleted after a period of time during the fill window. If there are titles being stored on an OCA that are no longer status, other OCAs can then fill from them, and so on. This process continues As the sun moves west and more members begin streaming, the fill window in this When the second tier of OCAs complete their download, they report back their

# 2. Tier fill: Available OCAs outside the manifest cluster configuration 1. Peer fill: Available OCAs within the same manifest cluster or the same subnet

## 3. Cache fill: Direct download from S3

## Example Scenario

given the option to fill from the fill master. communicate with the control plane to request a fill source for this title, they are control plane that it now has the title stored. The next time the other OCAs After the fill master OCA has completed its S3 download, it reports back to the



In the Netflix architecture diagram, a terminal can act as the first server, where the other OCAs can have the option of downloading from the "fill master OCA."	wherein the first server is a terminal adapted to act as relay server.	[126]
See above.	The method as claimed in claim 9,	[12a]
The control panel on the main server observes performance data to determine which "second tier" OCAs to complete their downloads with from a master OCA, to "report back their (second tier OCAs) status" to the main server, "and so on" to send transport requests to each new "master" OCA and new "second tier" OCA's, "and so on."		
Netflix heavily monitors performance and response times for OCAs in a network, as described above. This is so closely monitored by Netflix that it issues a "Route Performance Report" to its clients regarding performance of OCAs.		
Netflix' main server is adapted to send <b>transport requests</b> direct to at least one first target terminal on the basis of said terminal performance information. These transport requests are the commands sent by Netflix when its uses "appliance selection criteria" and commands where the "control pane steers clients to the best available OCAs." Netflix' wrote that "We steer clients to our OCA's based on an ISP's BCP advertisements, coupled with the routing and steering algorithms" This steering by Netflix is also included in transport request.		

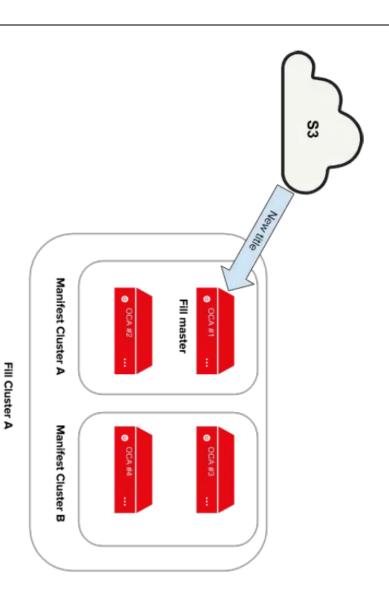
# 1. Peer fill: Available OCAs within the same manifest cluster or the same subnet

2. Tier fill: Available OCAs outside the manifest cluster configuration

- Cache fill: Direct download from S3

communicate with the control plane to request a fill source for this title, they are control plane that it now has the title stored. The next time the other OCAs After the fill master OCA has completed its S3 download, it reports back to the Example Scenario

given the option to fill from the fill master.



# Interaction with Client Devices and Netflix AWS Services

data). Essentially, OCA servers only do the following two things: OCAs do not store client data (for example - viewing history, DRM info, or member

- learned from the BGP peer (router or switch) they have a configured BGP Services (AWS). For example, they report health metrics, BGP routes they have Report their status to the Open Connect control plane services in Amazon Web
- Serve content via HTTP/HTTPS when it is requested by a client device

session with, and what files they have stored on disk.

clients via URL to the most optimal OCAs given their file availability, health, and network The control plane services in AWS take the data that the OCAs report and use it to steer

and handle the storage and interpretation of relevant telemetry about the playback files to OCAs nightly), compute optimal behavior for such things as file storage/hashing proximity to the client. The control plane services also control fill behavior (adding new experience

See Netflix Open Connect Overview, 4-5

acknowledgement of packets when content is transmitted: TCP data communication protocol, which most of the Internet included Netflix uses, requires an

transmission. This requires the receiver to respond with an acknowledgement message as it TCP achieves this using a technique known as positive acknowledgement with reidentical and in the same order as those sent. Since packet transfer by many networks is not reliable TCP is a reliable byte stream delivery service which guarantees that all bytes received will be receives the data. The sender keeps a record of each packet it sends and maintains a timer from

https://openconnect.zendesk.com/hc/en-us/articles/360035533071#routi	
<ul> <li>Traffic to OCA: Allow all destination addresses and ports.</li> <li>Traffic to OCA: Allow TCP 22, 53, 80, 179, 443, UDP 53 and 123 (source and destination), ICMP types 0, 3, 8, 11, and all ICMPv6 from any public IP/port. Allow all return traffic from any appliance-initiated connection (TCP established).</li> </ul>	
Netflix requires network traffic to OCA be in TCP protocol:	
Comer, Douglas E. (2006). Internetworking with TCP/IP: Principles, Protocols, and Architecture. Vol. 1 (5th ed.). Prentice Hall. ISBN 978-0-13-187671-2	
when the packet was sent. The sender re-transmits a packet if the timer expires before receiving th acknowledgement. The timer is needed in case a packet gets lost or corrupted.[	

#### 37

## Glossary of terms

Compound Annual Growth Rate

A temporary local copy of information that originated elsewhere. Thus for CDNs, a copy of files to be delivered to consumers, stored in a local server

## CDN (Content Delivery Network)

A distributed system of servers, designed to enable the efficient and reliable distribution of content over the internet

## CP (content provider)

An internet business whose focus is delivering content (rather than - say - e-commerce) to consumers. Netflix, CNN and YouTube are examples

across multiple servers

One step in a packet of data's journey

Fixed broadband

ISP (Internet Service Provider)

A company providing internet connectivity

## IXP (Internet Exchange Point)

to consumers (consumers or businesses). May provide fixed broadband, mobile data or both

A location where many networks meet to exchange traffic, avoiding the need for multiple bilateral connections

#### attency

The lag between a packet of data being sent and reaching its destination

## OTT (Over The Top)

Describes services delivered over another

network without being integrated with it.
YouTube, Facebook and Netflix are examples
(since they are not provided by telcos operating broadband networks)

## Packet loss

When a router is sent more data than it can handle, it discards a certain amount of data. This is known as packet loss. Typically the data in question will then be requested again from the source server

### Server

appropriate link

of data and sends them onwards down the

A switch on the internet, that receives packets

A computer that stores and transmits content

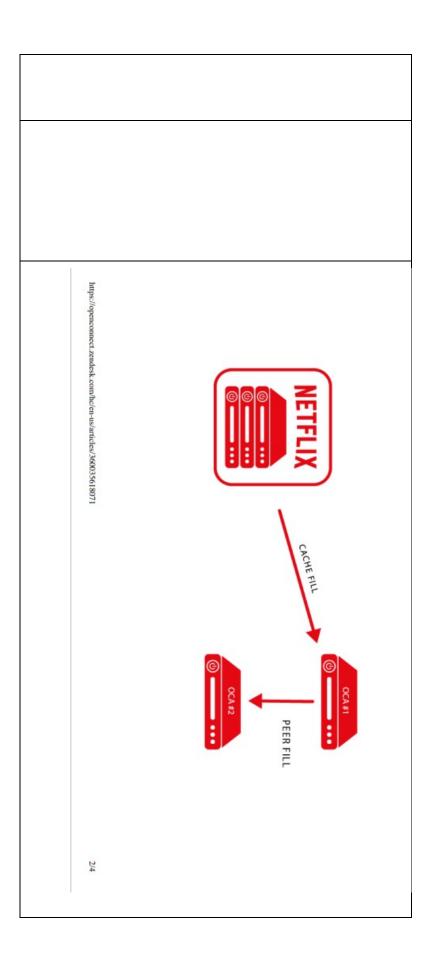
#### C#/=

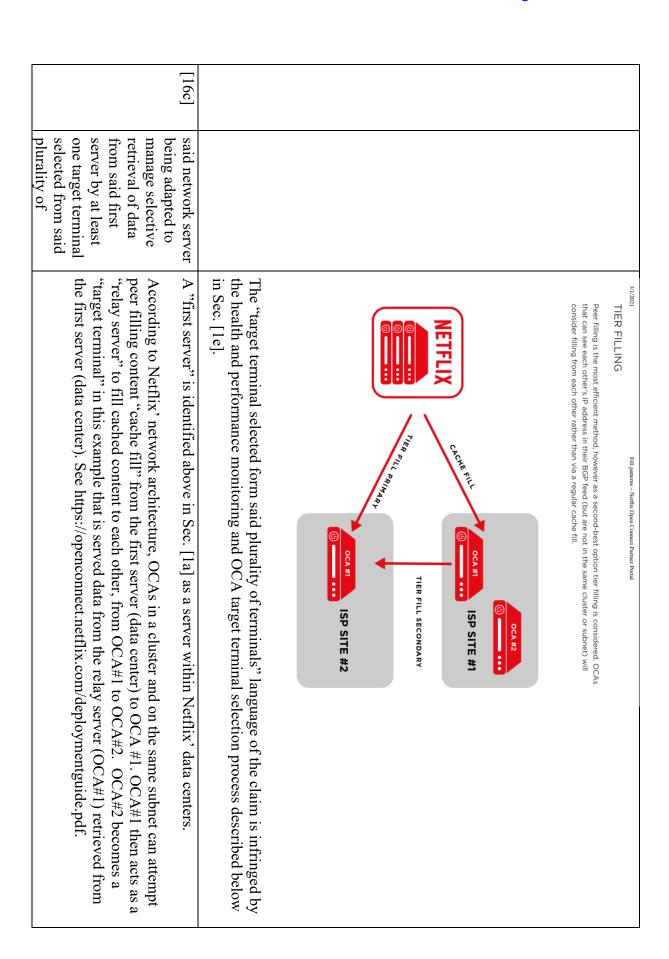
Transmission control protocol and internet protocol. The two foundational standards for data transmission that underpin the internet

Netflix runs the cache fill operation of Open Connect from a Netflix CCS (main server) that is hosted in AWS. See Open Connect Overview, p. 4-5.	network including:	
The building blocks of Open Connect are our suite of purpose-built <b>server appliances</b> , <b>called Open Connect Appliances (OCAs)</b> . <i>See</i> Open Connect Overview, p. 2. These are deployed directly inside ISP networks. Netflix provides the server hardware.	communication network, the data communication	
Netflix uses a system called <b>Open Connect</b> to deliver Netflix TV shows and movies to members world-wide. Netflix runs the operation of Open Connect from a Netflix application (CCS server) that is hosted in AWS. <i>See</i> Open Connect Overview, p. 4-5.	A network server adapted to operate as a main server in	[16a]
Netflix states, on its public technical blog, that binary format data can be used in image file formats sent through the network. <i>See</i> <a href="https://netflixtechblog.com/enhancing-the-netflix-ui-experience-with-hdr-1e7506ad3e8">https://netflixtechblog.com/enhancing-the-netflix-ui-experience-with-hdr-1e7506ad3e8</a>	including transmitting said data in binary format.	[15b]
See above.	The method as claimed in claim 9,	[15a]
In another example, "After <b>the fill master OCA</b> has completed its S3 download, <b>it reports back to the control plane</b> that it now has the title stored." See https://netflixtechblog.com/netflix-and-fill-c43a32b490c0		
Netflix uses TCP for internet streaming to send packets of data for video. Additionally, Netflix specifically looks at the number of TCP connections to determine internet speeds in accordance with testing of OCAs.		
Where does Netflix use TCP?		
"A cooperative approach to content delivery," Netflix (2021), 37.		

in the network information database, said network server being distinct from said first server.	m said o at ther hal minal stored		retrieved by at least one target terminal one target terminal from among said plurality of terminals, at least terminals at least terminals terminals at least terminal in deployment of Open Connect, Netflix provides Internet Service Providers with an OCA appliance direct "settlement-free interconnection (SFI)." The terminal OCA can "Connect Application of Open Connect (PNI) or IXP-based SFI peering to Netflix Open Connect Application of Open Connect (PNI) or IXP-based SFI peering to Netflix Open Connect Application of Open Connect (PNI) or IXP-based SFI peering to Netflix Open Connect Application of Open Connect (PNI) or IXP-based SFI peering to Netflix Open Connect Application of Open Connect (PNI) or IXP-based SFI peering to Netflix Open Connect Application of Open Connect Applicatio	d a first h from	[16b] a plurality of Netflix runs the operation of Open Connect from a terminals, a AWS. See Open Connect Overview, p. 4-5.
first server in "our data centers."  OCAs in a cluster and on the same subnet can attempt peer filling from each other. There is also Tier filling where if in different ISP sites. <a href="deeploymentguide.pdf">deploymentguide.pdf</a> (netflix.com)	ISPs who do not currently participate in pubic peering might want to consider that a single IX port can support multiple peering sessions, providing <b>direct access to various content</b> , cloud, and network providers.  Welcome to Open Connect, p. 3. Dkt 39 at p. 48.  The following diagram also illustrates access from a target terminal (OCA #1, OCA #2) to a Netflix	"Netflix has the ability to interconnect at a number of global data center facilities and public Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any network at IXP locations where we are mutually present and we consider private interconnection as appropriate."	In deployment of Open Connect, Netflix provides Internet Service Providers with an OCA appliance direct "settlement-free interconnection (SFI)." The terminal OCA can "Connect via direct Private Network Interconnect (PNI) or IXP-based SFI peering to Netflix Open Connect Appliances in our data centers."	In its global network, Netflix provides <b>data centers</b> such as an "S3" server (" <b>first server</b> ") housing <b>content</b> (" <b>data</b> ") <b>on at lease one server</b> , and provides OCA users (such as ISP's) direct access to these data centers over the Internet that are housing the content. One or more of these <b>data centers</b> house a " <b>first server</b> " according to the claims.	Netflix runs the operation of Open Connect from a Netflix application (CCS server) that is hosted in AWS. <i>See</i> Open Connect Overview, p. 4-5.

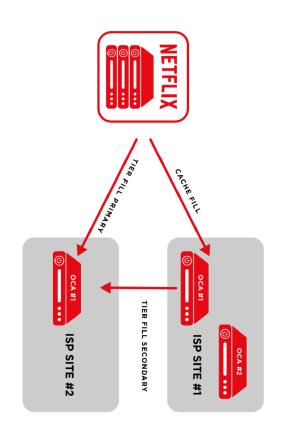






which then acts as a "relay server" to fill a target terminal (OCA#2) using "tier filling." In another example, Netflix's first server (data center) can cache fill content to a terminal OCA#1,

cluster or subnet) will consider filling from each other rather than via a regular cache fill." considered. OCAs that can see each other's IP address in their BGP feed (but are not in the same "Peer filling is the most efficient method, however as a second-best option tier filling is



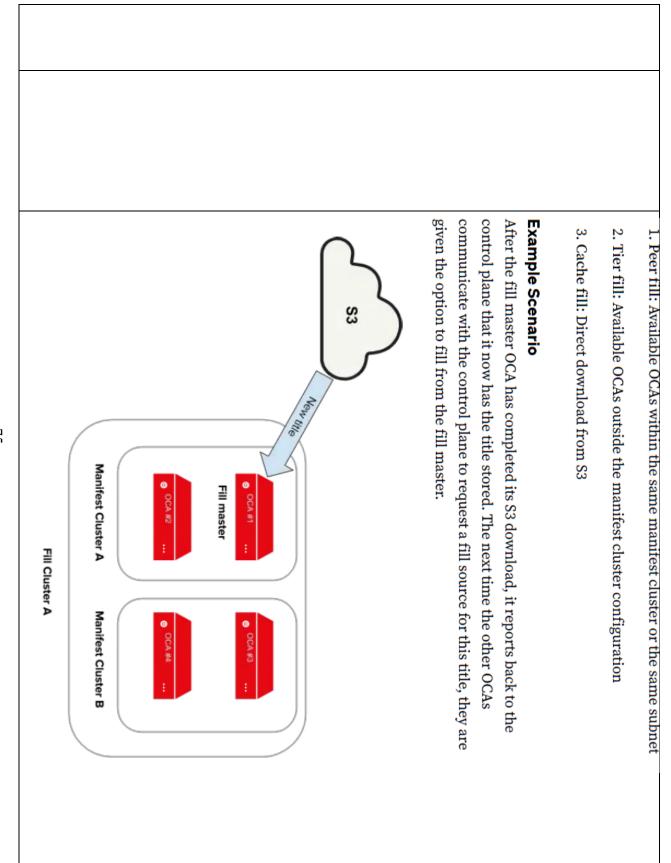
See https://openconnect.netflix.com/deploymentguide.pdf

servers, or "masters" that target terminals can use to gain, or fill, content: "The control plane elects the specified number of OCAs as masters...." OCAs can act as relay In another example, https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states that there are

The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to
3. Whether the OCA can go to S3, and how long it should wait before doing so
2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so
<ol> <li>How many hops away an OCA can go to download content, and how long it should wait before doing so</li> </ol>
A fill escalation policy defines:
• Fill escalation policies
• Fill master (number per fill cluster)
<ul> <li>BGP path attributes and physical location (latitude / longitude)</li> </ul>
To calculate the least expensive fill source, we take into account network state and some configuration parameters for each OCA that are set by the Open Connect Operations team. For example:
It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to <i>all</i> of our OCAs, so we use a tiered approach. The goal is to ensure that the title is passed from one part of our network to another using the most efficient route possible.
Calculating the Least Expensive Fill Source
• A calculated <i>route</i> cost — Described in the next section.
• Fill health — Can the fill source take on additional fill traffic?
<ul> <li>Title (content) availability — Does the fill source have the requested title stored:</li> </ul>

during its configured fill window, that includes a list of the new or updated titles	
in the manifest and what they are currently storing, each OCA will send a request,	
should be storing and serving to members. If there is a delta between the list of titles	
requesting (among other things) a manifest file that contains the list of titles they	
plane. OCAs communicate at regular intervals with the control plane services,	
popularity, etc. All of this information is aggregated and stored in the AWS control	
OCAs do not store any information about other OCAs in the network, title	
Fill Source Manifests	
The CCS server monitors the "fill health" of each potential download locations, or fill sources that includes OCAs, and uses a fill escalation policy based upon response times to determine if the performance of the download location can take on additional traffic or not and how long the OCA should wait for a response before escalation to the next download locations.	

serve this new title have it stored. time zones — until enough of the OCAs in our global network that need to be able to time zone ends, and the fill pattern continues as the fill window moves across other that ensures we don't interrupt any live sessions. needed, they are put into a delete manifest and then deleted after a period of time during the fill window. If there are titles being stored on an OCA that are no longer status, other OCAs can then fill from them, and so on. This process continues As the sun moves west and more members begin streaming, the fill window in this When the second tier of OCAs complete their download, they report back their



								information database;	ored
The CCS server monitors the "fill health" of each potential download locations, or fill sources that includes OCAs, and uses a fill escalation policy based upon response times to determine if the performance of the download location can take on additional traffic or not and how long the OCA should wait for a response before escalation to the next download locations. Thus, the manifest is assembled by the CCS server and download locations are assembled based on OCA performance	(Emphasis added.)	<ul> <li>10. How many hops away an OCA can go to download content, and how long it should wait before doing so</li> <li>11. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so</li> <li>12. Whether the OCA can go to S3, and how long it should wait before doing so</li> </ul>	A fill escalation policy defines:	A calculated route cost — Described in the next section.	Fill health — Can the fill source take on additional fill traffic?	Title (content) availability — Does the fill source have the requested title stored?	The response from the control plane in AWS is a ranked list of potential download locations, a.k.a. fill sources, for each title. The determination of the list takes into consideration several high-level factors:	The main server on CCS is further adapted to monitor response times of terminals in the network and in which terminals are selected to act as relay servers for a particular data transfers on the basis of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a> , Netflix wrote:	The control plane elects the specified number of OCAs as masters for a given title asset. <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a> Thus, the CCS server selects OCA terminals as a download location for a given title asset.

The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to reach farther with less delay in order to grab that content and then share it locally with non-masters.	3. Whether the OCA can go to S3, and how long it should wait before doing so	2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so	1. How many hops away an OCA can go to download content, and how long it should wait before doing so	It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to all of our OCAs, so we use a tiered approach A fill escalation policy defines:	The CCS server information location list provides the address of a first server, called the "S3" server, for a download location and provides other download locations of a second and additional master OCAs:	https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one second OCA can be selected based on their relative performance:	A "desired manifest" and when needed an "emergency manifest" are transport requests posted by Netflix to the CCS (Control Content Service) server in AWS. These manifests are data structures that indicate what titles need to be transferred to a specific OCA terminal, either during an off-peak hours download window, or urgently if there is an emergent need. these manifests are fetched by regularly from the CCS by each OCA terminal, which traverses the desired manifest or emergency manifest. OCAs then then query the CCS for location information files that list where each title on the desired manifest that is needed by an OCA can be downloaded.	information, and the manifest is sent to a given OCA. Therefore, the transport request is sent on a basis of said terminal performance information.

[16e]	
wherein terminals adapted to act as relay servers are adapted to modify transport requests received from said network server or from other relay servers and to transmit the modified transport request to selected target terminals	
OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle. The OCAs work in a network to distribute updates among each other and to include further OCAs to which updates and content can be sent. <i>See</i> Open Connect Overview, p. 5; Fill Patterns, pp. 1-3.  Netflix' OCA that are adapted to act as relay servers (see 1d above) are adapted to modify transport requests received from the main server or from other relay servers and transmit the modified transport request to selected target terminals that includes addresses of further target terminals.  The CCS server will to order OCA terminals to peer or tier fill from using OCAs selected by the CCS server. The CCS server uses Appliance Section Criteria to select OCA terminals as targets in the to transfer titles from: 1) the terminal appliance that receives the most-specific route; 2) the terminal appliance that receives the route to the client's netblock with the shortest AS path; 3) the terminal appliance that receives the route to the client's netblock with the lowest multi-exit	As stated in [1e], the CCS monitors the "fill health" and performance of download locations, which is based on performance of the OCAs, to determine if that OCA will be selected as a download location or not.  This desired manifest and emergency manifest along with the downloaded location information files and fill policy for master OCAs and a second (target terminal) OCA to fill from a (first terminal) OCA on the CCS are equivalent the "each such transport request includes details of data to be retrieved, the address of the first server from which the data is to be requested by the first target terminal, the addresses of at least one second target terminal to which the data from the first server to be relayed by the first target terminal." The functions are substantially the same - to send a request from a main server (CCS) to a terminal (OCA) with instructions to download or "fill" content from specific master terminal/OCA addresses and includes an address of at least one second terminal/OCA. The way this is accomplished is substantially the same. The claim states that the main server is adapted to send transport requests that include download locations direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix posts a desired manifest or emergency manifest to the CCS server along with download locations that are intended for each OCA to read on a regular basis. The results are substantially the same - to transmit commands from the main server/CCS server to a terminal/OCA to download content data in the manifest that needs to be retrieved from various addresses.

relay server; and request is to act as modified transport the recipient of the further target addresses of transport request the modified request, wherein in the transport terminals for which further includes locations of master OCAs for each individual title needed by an OCA to fill its delta:

of these master or target terminals in the desired manifest, which is loaded by an OCA terminal in discriminator; 4) the geographically closest appliance. The CCS server includes the URL addresses its memory or hard drive space in order to select an OCA for downloading titles from

terminals identified

on the delta. The CCS responds, as stated in [1f] with a list of URLs that are downloadable "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing The OCA terminal will then query the CCS terminal for a list of download locations for each title from the actual manifest are termed the "delta" or difference between actual and desired manifest. After receiving the desired manifest, an OCA terminal will compare the manifest with its own

updated titles that it needs. The response from the control plane in AWS is a ranked list of things) a manifest file that contains the list of titles they should be storing and serving to members https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 potential download locations, a.k.a. fill sources, for each title." (Emphasis added.) See If there is a delta between the list of titles in the manifest and what they are currently storing, each "OCAs communicate at regular intervals with the control plane services, requesting (among other OCA will send a request, during its configured fill window, that includes a list of the new or

and then requesting a delta-listed title from the list of a master OCAs, is equivalent to a modified desired manifest) to request a title or titles from further target terminals, or master OCAs. transport request. The OCA is using a modified list of titles (delta or missing titles list from the The action of an OCA requesting download locations (master OCAs) for its delta list from the CCS

request, wherein the modified transport request further includes addresses of further target terminals transport request to selected target terminals from a set of target terminals identified in the transport requests received from the main server or from other relay servers and to transmit the modified This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport for which the recipient of the modified transport request is to act as relay server."

substantially the same. A delta is a modification of the manifest list, or in other words a subset list with URLs of those relay servers or URLs of master OCAs. The way the actions are performed are relay servers or OCAs a modification of the original transport request, or desired manifest, along The function of the two actions are substantially the same, which is to transmit to other selected

multiple delta titles to selected OCAs) are sent to addresses (URLs) of further selected relay servers original transport request (desired manifest) is transmitted to selected relay terminals (single or substantially similar: further relay terminal addresses are sent to the OCAs, a modification of the of URL locations to download the modified list titles of its delta list, after which the delta list is of what it is supposed to download. Instead of the terminal, or OCA, transmitting the delta list to then sent to another OCA via a series of modified transfer requests. The result of these actions are another OCA, an "actual manifest" is sent to the CCS server which responds to the OCA with a list (OCAs) in the form of a request for one or more titles from one or more master OCAs.

clustered or if they are in the same subnet, will attempt to peer or tier fill from each other https://openconnect.zendesk.com/hc/en-us/articles/360035618071-Fill-patterns Alternatively and equivalently, Netflix documentation discloses that OCA terminals, if they are

geographically closest terminal. The OCA terminal includes the URL addresses of these terminals titles from: 1) the terminal appliance that receives the most-specific route; 2) the appliance terminal save this information. In general, appliances determine where to receive fill using selection criteria Patterns, pp. 1-4. in its memory or hard drive space in order to select an OCA for downloading titles from. See Fill that receives the route to the client's netblock with the lowest multi-exit discriminator; 4) the that receives the route to the client's netblock with the shortest AS path; 3) the appliance terminal that is used by Netflix client devices. The OCA terminals then use a similar Appliance Section Criteria as the CCS server uses to select OCA terminals as targets in the subnet or cluster to transfer OCA terminals in a subnet or cluster broadcast their IP and physical locations to one another and

requests for each title using the URL of the target OCA. request the delta list items to the selected OCA(s) in the subnet or cluster in the form of download After selecting an OCA master using the selection criteria, the OCA terminal will transmit and from the actual manifest are termed the "delta" or difference between actual and desired manifest "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing After receiving the desired manifest, an OCA terminal will compare the manifest with its own

transport request to selected target terminals from a set of target terminals identified in the transport requests received from the main server or from other relay servers and to transmit the modified This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport

					[16f]	
	acknowledge receipt of the last packet of a series routed thereto.	directly with said main server to	said target terminals and each of said terminals are adapted to communicate	divided into a series of packets for transmission to	wherein data to be retrieved by said	
TCP is a reliable byte stream delivery service which guarantees that all bytes received will be identical and in the same order as those sent. Since packet transfer by many networks is not reliable,	TCP data communication protocol, which most of the Internet included Netflix uses, requires an acknowledgement of packets when content is transmitted:	There is also evidence that the cache fill transfer uses TCP/IP protocols for packetized data transfer.	Additionally, the last packet of a download for a cache fill is associated with a notice to the CCS that the download has been completed. The system can also use parallel processes to perform cache transfer using TCP/IP protocols. This is only possible if the CCS is working on a different section of the file at the same time, which is packetized file transfer.	The Open Connect network can make parallelized cache fill transfers. As such, it must be transferring "packets" of the file. There is evidence that each file is downloaded for cache fill in packet sizes up to 16 kilobytes.	Netflix' OCAs are adapted to communicate with the main server which is hosted by AWS. The OCA's "Report their status to the Open Connect control panel services in Amazon Web Services":	request, wherein the modified transport request further includes addresses of further target terminals for which the recipient of the modified transport request is to act as relay server."  The function of the two actions are substantially the same, which is to transmit to other selected relay servers or OCAs a modified list of the original transport request or manifest. The way the actions are performed are substantially the same. A modified desired manifest list is a subset list called the "delta." A relay server (OCA) creates a modified transport request (delta list) of its missing titles and transports this list to another relay server (OCA sends requests to other selected OCAs for the titles on the delta list in a series of requests) and includes addresses of the other relay servers (URLs of the selected OCA terminals in the request as an address to the selected OCAs). The result of these actions are substantially similar: a modification of the original transport request (desired manifest) is transmitted to selected relay terminals (selected OCAs) with addresses (URLs) of the selected relay servers (OCAs).

https://openconnect.zendesk.com/hc/en-us/articles/360035533071#routi	<ul> <li>Traffic from OCA: Allow all destination addresses and ports.</li> <li>Traffic to OCA: Allow TCP 22, 53, 80, 179, 443, UDP 53 and 123 (source and destination), ICMP types 0, 3, 8, 11, and all ICMPv6 from any public IP/port. Allow all return traffic from any appliance-initiated connection (TCP established).</li> </ul>	Netflix requires network traffic to OCA be in TCP protocol:	Comer, Douglas E. (2006). Internetworking with TCP/IP: Principles, Protocols, and Architecture. Vol. 1 (5th ed.). Prentice Hall. ISBN 978-0-13-187671-2	TCP achieves this using a technique known as positive acknowledgement with retransmission. This requires the receiver to respond with an acknowledgement message as it receives the data. The sender keeps a record of each packet it sends and maintains a timer from when the packet was sent. The sender re-transmits a packet if the timer expires before receiving the acknowledgement. The timer is needed in case a packet gets lost or corrupted.

# Glossary of terms

A temporary local copy of information that in a local server of files to be delivered to consumers, stored originated elsewhere. Thus for CDNs, a copy

# **CDN (Content Delivery Network)**

content over the internet enable the efficient and reliable distribution of A distributed system of servers, designed to

## CP (content provider)

content (rather than - say - e-commerce) An internet business whose focus is delivering to consumers. Netflix, CNN and YouTube

from the source server

## Server

of data and sends them onwards down the appropriate link A switch on the internet, that receives packets

## across multiple servers

One step in a packet of data's journey

ISP (Internet Service Provider)

provide fixed broadband, mobile data or both to consumers (consumers or businesses). May A company providing internet connectivity

IXP (Internet Exchange Point)

to exchange traffic, avoiding the need A location where many networks meet

for multiple bilateral connections

## TCP/IP

protocol. The two foundational standards for Transmission control protocol and internet

Compound Annual Growth Rate

sent and reaching its destination

The lag between a packet of data being

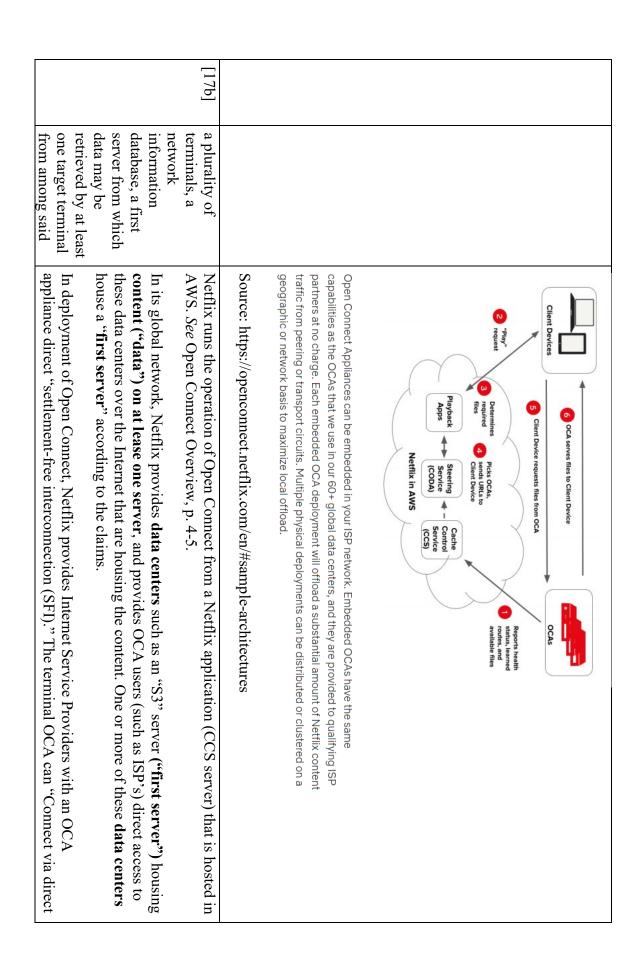
Packet loss broadband networks) (since they are not provided by telcos operating YouTube, Facebook and Netflix are examples network without being integrated with it. Describes services delivered over another OTT (Over The Top)

data in question will then be requested again This is known as packet loss. Typically the handle, it discards a certain amount of data When a router is sent more data than it can

Fixed broadband

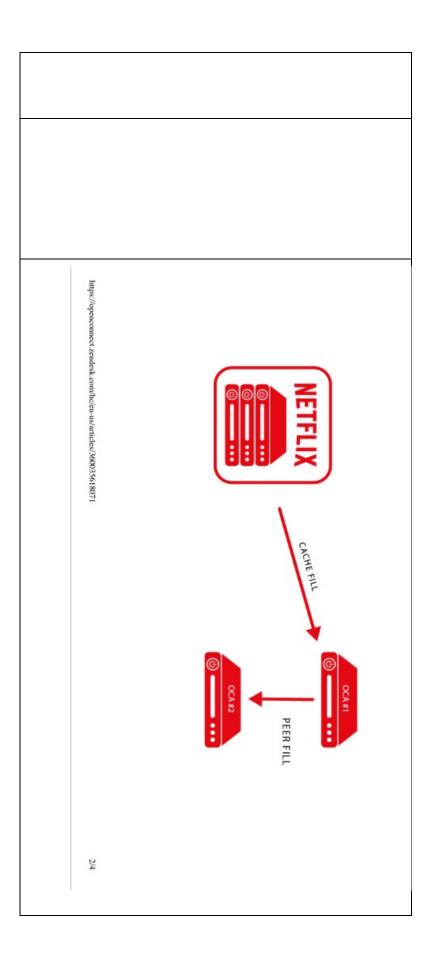
transmits content A computer that stores and

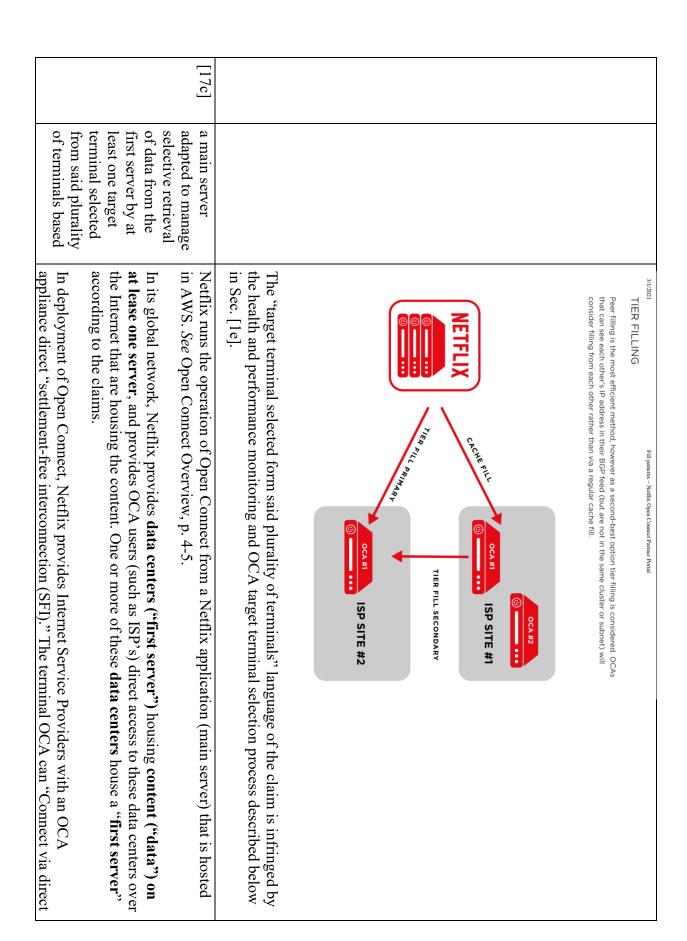
al Netflix uses a system called <b>Open Connect</b> to deliver Netflix TV shows and movies to members world-wide.  The building blocks of Open Connect are our suite of purpose-built <b>server appliances, called Open Connect Appliances (OCAs)</b> . <i>See</i> Open Connect Overview, p. 2. These are deployed directly inside ISP networks. Netflix provides the server hardware. The OCAs report to a Open Connect control plane (CCS server) to control fill behavior (adding new files to OCAs nightly) and to compute and/or store data. <i>See id.</i> p. 3-4. Accordingly, OCAs include both an input mechanism and display mechanism.	A network terminal to operate as a relay server in a data communication network, the data communication network including:	[17a]
In another example, "After <b>the fill master OCA</b> has completed its S3 download, <b>it reports back to the control plane</b> that it now has the title stored." See https://netflixtechblog.com/netflix-and-fill-c43a32b490c0		
Netflix uses TCP for internet streaming to send packets of data for video. Additionally, Netflix specifically looks at the number of TCP connections to determine internet speeds in accordance with testing of OCAs.		
Where does Netflix use TCP?		
"A cooperative approach to content delivery," Netflix (2021), 37.		



Tier filling where if in different ISP sites. deploymentguide.pdf (netflix.com)	
OCAs in a cluster and on the same subnet can attempt neer filling from each other. There is also	
The following diagram also illustrates access from a target terminal (OCA $\#1$ , OCA $\#2$ ) to a Netflix first server in "our data centers."	
Welcome to Open Connect, p. 3. Dkt 39 at p. 48.	
ISPs who do not currently participate in pubic peering might want to consider that a single IX port can support multiple peering sessions, providing <b>direct access to various content</b> , cloud, and network providers.	
"Netflix has the ability to interconnect at a number of global data center facilities and public Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any network at IXP locations where we are mutually present and we consider private interconnection as appropriate."	
Private Network Interconnect (PNI) or IXP-based SFI peering to Netflix Open Connect Appliances in our data centers."	plurality of terminals; and







are selected to act as relay servers for a particular data transfers on the basis of their relative response times;	server is further adapted to monitor response times of terminals in the network and in which terminals	on terminal performance data stored in the network information database, and wherein the main
The following diagram also illustrates access from a target terminal (OCA #1, OCA #2) to a Netflix first server in "our data centers."	ISPs who do not currently participate in pubic peering might want to consider that a single IX port can support multiple peering sessions, providing <b>direct access to various content</b> , cloud, and network providers.  Welcome to Open Connect, p. 3. Dkt 39 at p. 48.	in our data centers."  "Netflix has the ability to interconnect at a number of global data center facilities and public Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any network at IXP locations where we are mutually present and we consider private interconnection as appropriate."

## consider filling from each other rather than via a regular cache fill. that can see each other's IP address in their BGP feed (but are not in the same cluster or subnet) will Peer filling is the most efficient method, however as a second-best option tier filling is considered. OCAs TIER FILLING OCA #1 TIER FILL SECONDARY **ISP SITE #2 ISP SITE #1**

"The control plane elects the specified number of OCAs as masters..." OCAs can act as relay the health monitoring and OCA target terminal selection process described below in Sec. [1e]. The "target terminal selected form said plurality of terminals" language of the claim is infringed by In another example, https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states that there are

servers, or "masters" that target terminals can use to gain, or fill, content:

The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to	The control pla asset. The fill e
3. Whether the OCA can go to S3, and how long it should wait before doing so	3. Whether th
<ol><li>Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so</li></ol>	2. Whether th defined abo
<ol> <li>How many hops away an OCA can go to download content, and how long it should wait before doing so</li> </ol>	1. How many should wait
olicy defines:	A fill escalation policy defines:
1 policies	• Fill escalation policies
Fill master (number per fill cluster)	• Fill master
BGP path attributes and physical location (latitude / longitude)	BGP path at
To calculate the least expensive fill source, we take into account network state and some configuration parameters for each OCA that are set by the Open Connect Operations team. For example:	To calculate the least expensive some configuration parameter: Operations team. For example:
Calculating the Least Expensive Fill Source It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to <i>all</i> of our OCAs, so we use a tiered approach. The goal is to ensure that the title is passed from one part of our network to another using the most efficient route possible.	Calculating the It would be inear from S3 to <i>all</i> of the title is passe route possible.
A calculated <i>route</i> cost — Described in the next section.	• A calculated
Fill health — Can the fill source take on additional fill traffic?	• Fill health –
Title (content) availability — Does the fill source have the requested title stored:	• Title (conte

[17d]and an indication network terminal relayed by the which the data target terminal to network terminal, requested by the the data is to be server from which address of the first retrieved, the details of data to be request including each such transport network terminal, sent to said responding to receiving and retrieved from said serving data relay server for adapted to act as said network first server is to be retrieved from the least one second the addresses of at transport requests terminal by first server to at terminal being least one target

each other - minimizing significant usage of internet "backbone" capacity during the update network during these configurable time windows. OCAs can also download updates from use of non-peak bandwidth to download the vast majority of content updates to the OCAs in All OCA deployments are constantly monitored to ensure reliability and efficiency. Netflix makes

an emergent need a specific OCA terminal, either during an off-peak hours download window, or urgently if there is server in AWS. These manifests are data structures that indicate what titles need to be transferred to A "desired manifest" and when needed an "emergency manifest" are posted by Netflix to the CCS

each OCA terminal is mandated to check regularly for manifests and download the manifests to the manifest. The CCS server acts as an email inbox for manifests sent directly to the OCAs, where regularly from the CCS by each OCA terminal, which traverses the desired manifest or emergency window, or urgently if there is an emergent need. Equivalently, these manifests are fetched by titles need to be transferred to a specific OCA terminal, either during an off-peak hours download A "desired manifest" and when needed an "emergency manifest" are posted by Netflix to the CCS local OCA terminal hard drive (Control Content Service) server in AWS. These manifests are data structures that indicate what

server/CCS server to a terminal/OCA to download content data in the manifest that needs to be posts a desired manifest or emergency manifest to the CCS server that is intended for each OCA in direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix substantially the same. The claim states that the main server is adapted to send transport requests a terminal (OCA) with instructions to download or "fill" content. The way this is accomplished is equivalent the "server is adapted to send transport requests direct to at least one first target the network. The results are substantially the same - to transmit commands from the main This desired manifest and emergency manifest and the download location data on the CCS terminal." The functions are substantially the same - to send a request from a main server (CCS) to

Netflix to the CCS (Control Content Service) server in AWS. These manifests are data structures that indicate what titles need to be transferred to a specific OCA terminal, either during an off-peak A "desired manifest" and when needed an "emergency manifest" are transport requests posted by

of a relative

							database	performance in the netwo information	terminal bas the terminal	performance further target
								e stored ork	sed on	ofa
This desired manifest and emergency manifest along with the downloaded location information files and fill policy for master OCAs and a second (target terminal) OCA to fill from a (first terminal) OCA on the CCS are equivalent the "each such transport request includes details of data to be retrieved, the address of the first server from which the data is to be requested by the first	As stated in [1e], the CCS monitors the "fill health" and performance of download locations, which is based on performance of the OCAs, to determine if that OCA will be selected as a download location or not.	The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to reach farther with less delay in order to grab that content and then share it locally with non-masters.	3. Whether the OCA can go to S3, and how long it should wait before doing so	2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so	1. How many hops away an OCA can go to download content, and how long it should wait before doing so	It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to all of our OCAs, so we use a tiered approach A fill escalation policy defines:	The CCS server information location list provides the address of a first server, called the "S3" server, for a download location and provides other download locations of a second and additional master OCAs:	https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one second OCA can be selected based on their relative performance:	manifest. OCAs then then query the CCS for location information files that list where each title on the desired manifest that is needed by an OCA can be downloaded.	hours download window, or urgently if there is an emergent need. these manifests are fetched by regularly from the CCS by each OCA terminal, which traverses the desired manifest or emergency

manifest that needs to be retrieved from various addresses. commands from the main server/CCS server to a terminal/OCA to download content data in the manifest or emergency manifest to the CCS server along with download locations that are intended one first target terminal over a network, and on the Open Connect Network, Netflix posts a desired main server is adapted to send transport requests that include download locations direct to at least content from specific master terminal/OCA addresses and includes an address of at least one second a request from a main server (CCS) to a terminal (OCA) with instructions to download or "fill" server to be relayed by the first target terminal." The functions are substantially the same - to send terminal/OCA. The way this is accomplished is substantially the same. The claim states that the for each OCA to read on a regular basis. The results are substantially the same - to transmit target terminal, the addresses of at least one second target terminal to which the data from the first

and in which terminals are selected to act as relay servers for a particular data transfers on the basis The main server on CCS is further adapted to monitor response times of terminals in the network terminals as a download location for a given title asset The control plane elects the specified number of OCAs as masters for a given title asset https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 Thus, the CCS server selects OCA

of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a> fill sources, for each title. The determination of the list takes into consideration several high-level Netflix wrote: The response from the control plane in AWS is a ranked list of potential download locations, a.k.a

Title (content) availability — Does the fill source have the requested title stored?

Fill health — Can the fill source take on additional fill traffic?

A calculated route cost — Described in the next section.

A fill escalation policy defines:

relay server; and request is to act as modified transport further target addresses of further includes the recipient of the terminals for which transport request

the modified "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing locations of master OCAs for each individual title needed by an OCA to fill its delta: on the delta. The CCS responds, as stated in [1f] with a list of URLs that are downloadable The OCA terminal will then query the CCS terminal for a list of download locations for each title from the actual manifest are termed the "delta" or difference between actual and desired manifest After receiving the desired manifest, an OCA terminal will compare the manifest with its own

If there is a **delta** between the list of titles in the manifest and what they are currently storing, each things) a manifest file that contains the list of titles they should be storing and serving to members "OCAs communicate at regular intervals with the control plane services, requesting (among other https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 updated titles that it needs. The response from the control plane in AWS is a ranked list of potential download locations, a.k.a. fill sources, for each title." (Emphasis added.) See OCA will send a request, during its configured fill window, that includes a list of the new or

desired manifest) to request a title or titles from further target terminals, or master OCAs. and then requesting a delta-listed title from the list of a master OCAs, is equivalent to a modified transport request. The OCA is using a modified list of titles (delta or missing titles list from the The action of an OCA requesting download locations (master OCAs) for its delta list from the CCS

requests received from the main server or from other relay servers and to transmit the modified request, wherein the modified transport request further includes addresses of further target terminals for which the recipient of the modified transport request is to act as relay server." transport request to selected target terminals from a set of target terminals identified in the transport This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport

of URL locations to download the modified list titles of its delta list, after which the delta list is of what it is supposed to download. Instead of the terminal, or OCA, transmitting the delta list to substantially the same. A delta is a modification of the manifest list, or in other words a subset list with URLs of those relay servers or URLs of master OCAs. The way the actions are performed are relay servers or OCAs a modification of the original transport request, or desired manifest, along another OCA, an "actual manifest" is sent to the CCS server which responds to the OCA with a list then sent to another OCA via a series of modified transfer requests. The result of these actions are The function of the two actions are substantially the same, which is to transmit to other selected

multiple delta titles to selected OCAs) are sent to addresses (URLs) of further selected relay servers original transport request (desired manifest) is transmitted to selected relay terminals (single or substantially similar: further relay terminal addresses are sent to the OCAs, a modification of the (OCAs) in the form of a request for one or more titles from one or more master OCAs

clustered or if they are in the same subnet, will attempt to peer or tier fill from each other. Alternatively and equivalently, Netflix documentation discloses that OCA terminals, if they are https://openconnect.zendesk.com/hc/en-us/articles/360035618071-Fill-patterns

geographically closest terminal. The OCA terminal includes the URL addresses of these terminals save this information. In general, appliances determine where to receive fill using selection criteria Patterns, pp. 1-4. in its memory or hard drive space in order to select an OCA for downloading titles from. See Fill that receives the route to the client's netblock with the lowest multi-exit discriminator; 4) the that receives the route to the client's netblock with the shortest AS path; 3) the appliance terminal titles from: 1) the terminal appliance that receives the most-specific route; 2) the appliance terminal that is used by Netflix client devices. The OCA terminals then use a similar Appliance Section Criteria as the CCS server uses to select OCA terminals as targets in the subnet or cluster to transfer OCA terminals in a subnet or cluster broadcast their IP and physical locations to one another and

After selecting an OCA master using the selection criteria, the OCA terminal will transmit and "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing requests for each title using the URL of the target OCA. request the delta list items to the selected OCA(s) in the subnet or cluster in the form of download from the actual manifest are termed the "delta" or difference between actual and desired manifest. After receiving the desired manifest, an OCA terminal will compare the manifest with its own

request, wherein the modified transport request further includes addresses of further target terminals transport request to selected target terminals from a set of target terminals identified in the transport requests received from the main server or from other relay servers and to transmit the modified for which the recipient of the modified transport request is to act as relay server." This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport

https://openconnect.zendesk.com/hc/en-us/articles/360035533071#routi	
• Traffic to OCA: Allow TCP 22, 53, 80, 179, 443, UDP 53 and 123 (source and destination), ICMP types 0, 3, 8, 11, and all ICMPv6 from any public IP/port. Allow all return traffic from any appliance-initiated connection (TCP established).	
Traffic from OCA: Allow all destination addresses and ports.	
Netflix requires network traffic to OCA be in TCP protocol:	
Comer, Douglas E. (2006). Internetworking with TCP/IP: Principles, Protocols, and Architecture. Vol. 1 (5th ed.). Prentice Hall. ISBN 978-0-13-187671-2	
when the packet was sent. The sender re-transmits a packet if the timer expires before receiving the acknowledgement. The timer is needed in case a packet gets lost or corrupted.	

# Glossary of terms

Compound Annual Growth Rate

A temporary local copy of information that in a local server of files to be delivered to consumers, stored originated elsewhere. Thus for CDNs, a copy

# **CDN (Content Delivery Network)**

CP (content provider) content over the internet enable the efficient and reliable distribution of A distributed system of servers, designed to

content (rather than - say - e-commerce) An internet business whose focus is delivering to consumers. Netflix, CNN and YouTube

of data and sends them onwards down the A computer that stores and Server appropriate link A switch on the internet, that receives packets

## IXP (Internet Exchange Point)

provide fixed broadband, mobile data or both to consumers (consumers or businesses). May

to exchange traffic, avoiding the need A location where many networks meet for multiple bilateral connections

sent and reaching its destination The lag between a packet of data being

## OTT (Over The Top)

broadband networks) (since they are not provided by telcos operating YouTube, Facebook and Netflix are examples network without being integrated with it. Describes services delivered over another

## Packet loss

from the source server data in question will then be requested again This is known as packet loss. Typically the handle, it discards a certain amount of data When a router is sent more data than it can

## TCP/IP

transmits content

ISP (Internet Service Provider)

A company providing internet connectivity

One step in a packet of data's journey

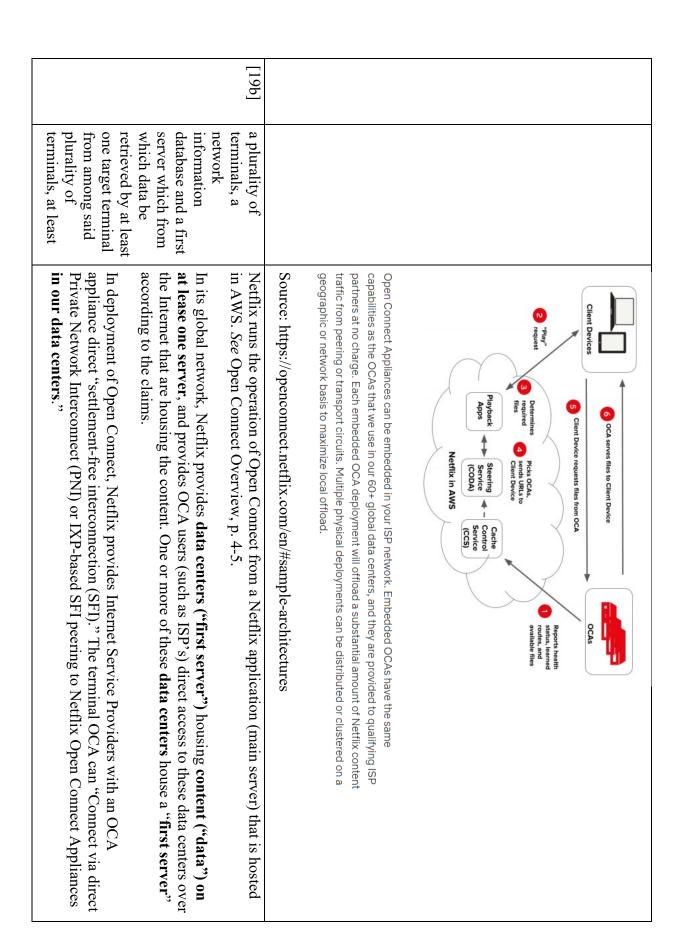
across multiple servers

횽

Fixed broadband

data transmission that underpin the internet protocol. The two foundational standards for Transmission control protocol and internet

[19a]	
A computer program product for enabling a network server to operate as a main server in a data communication network, the data communication network including:	
Netflix uses a system called <b>Open Connect</b> to deliver Netflix TV shows and movies to members world-wide.  The building blocks of Open Connect are our suite of purpose-built <b>server appliances, called Open Connect Appliances (OCAs)</b> . See Open Connect Overview, p. 2. These are deployed directly inside ISP networks. Netflix provides the server hardware. The OCAs report to a Open Connect control plane to control fill behavior (adding new files to OCAs nightly) and to compute and/or store data. See id. p. 3-4. Accordingly, OCAs include both an input mechanism and display mechanism  Netflix runs the operation of Open Connect from a Netflix application (CCS server) that is hosted in AWS. It is on information and belief that because Netflix Open Connect system includes the infringing elements and performs functions that infringe, there is the requisite software, or computer program, to carry out the functionality of Netflix's Open Connect system.	"A cooperative approach to content delivery," Netflix (2021), 37.  Where does Netflix use TCP?  Netflix uses TCP for internet streaming to send packets of data for video. Additionally, Netflix specifically looks at the number of TCP connections to determine internet speeds in accordance with testing of OCAs.  In another example, "After the fill master OCA has completed its S3 download, it reports back to the control plane that it now has the title stored." See https://netflixtechblog.com/netflix-and-fill-c43a32b490c0



Document 99-2

comprising:

computer program distinct from said server being database, said main information in the network information stored based on terminal serving data relay servers for adapted to act as first server, said pertormance target terminal retrieved from said least one further first server to at terminals being input required from our ISP partners. See Open Connect Overview, p. 5. network providers.

network at IXP locations where we are mutually present and we consider private interconnection as "Netflix has the ability to interconnect at a number of global data center facilities and public Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any

can support multiple peering sessions, providing direct access to various content, cloud, and ISPs who do not currently participate in pubic peering might want to consider that a single IX port

Welcome to Open Connect, p. 3. Dkt 39 at p. 48

first server in "our data centers." The following diagram also illustrates access from a target terminal (OCA #1, OCA #2) to a Netflix

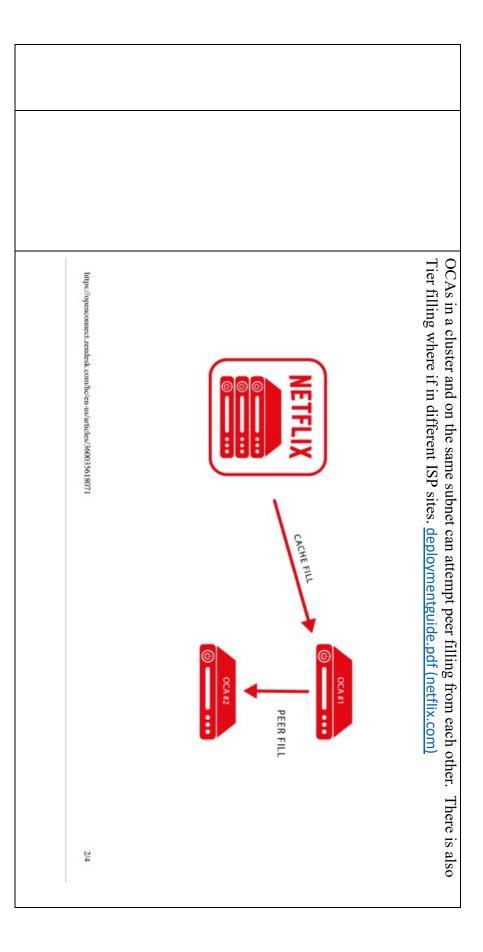
reliability and efficiency. We troubleshoot and proactively fix most issues remotely with minimal requirements evolve. See Open Connect Overview, p. 3. Netflix constantly measures and analyzes [OCA] performance and augment capacity as All OCA deployments are constantly monitored by the Open Connect Operations team to ensure

Additionally, OCAs periodically report health. *Id.* at 4.

Monitoring, Maintenance, and Updates

of our quality standards, we simply replace it - at no cost to our partners performance degrades to the point where a server is no longer functioning in the range status and performance, we provide a Partner Portal where they can do so. If hardware required from our ISP partners. If partners wish to monitor their own embedded OCAs' efficiency. We troubleshoot and proactively fix most issues remotely with minimal input constantly monitored by the Open Connect Operations team to ensure reliability and All of our OCA deployments, whether in IXPs or embedded in ISP networks, are

and in which terminals are selected to act as relay servers for a particular data transfers on the basis The main server on CCS is further adapted to monitor response times of terminals in the network Saving this information to a database is common practice within network management should wait for a response before escalation to the next download locations performance of the download location can take on additional traffic or not and how long the OCA A fill escalation policy defines: Netflix wrote: of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a> includes OCAs, and uses a fill escalation policy based upon response times to determine if the fill sources, for each title. The determination of the list takes into consideration several high-level The response from the control plane in AWS is a ranked list of potential download locations, a.k.a The CCS server monitors the "fill health" of each potential download locations, or fill sources that (Emphasis added.) 2 How many hops away an OCA can go to download content, and how long it should wait Whether the OCA can go to S3, and how long it should wait before doing so above), and how long it should wait before doing so Whether the OCA can go to the entire Open Connect network (beyond the hops defined before doing so Fill health — Can the fill source take on additional fill traffic? A calculated route cost — Described in the next section Title (content) availability — Does the fill source have the requested title stored?



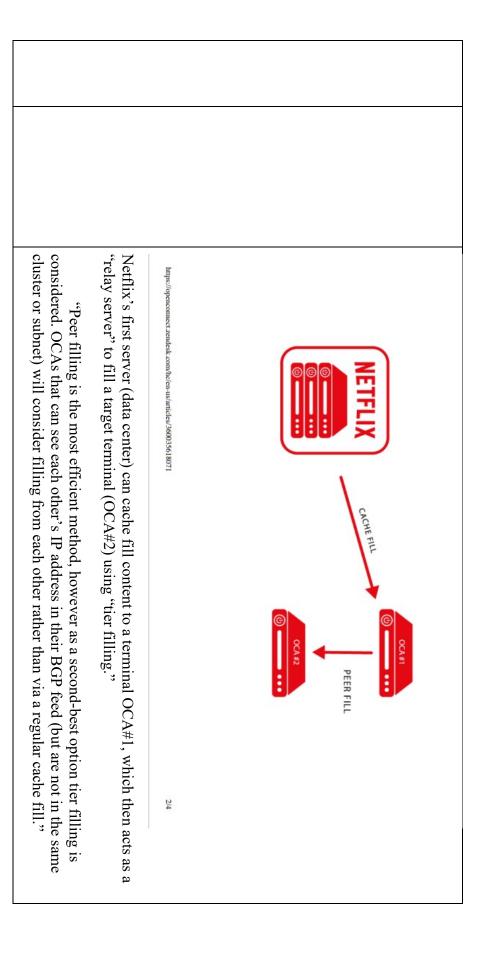
### consider filling from each other rather than via a regular cache fill. that can see each other's IP address in their BGP feed (but are not in the same cluster or subnet) will TIER FILLING Peer filling is the most efficient method, however as a second-best option tier filling is considered. OCAs OCA #1 TIER FILL SECONDARY **ISP SITE #2 ISP SITE #1**

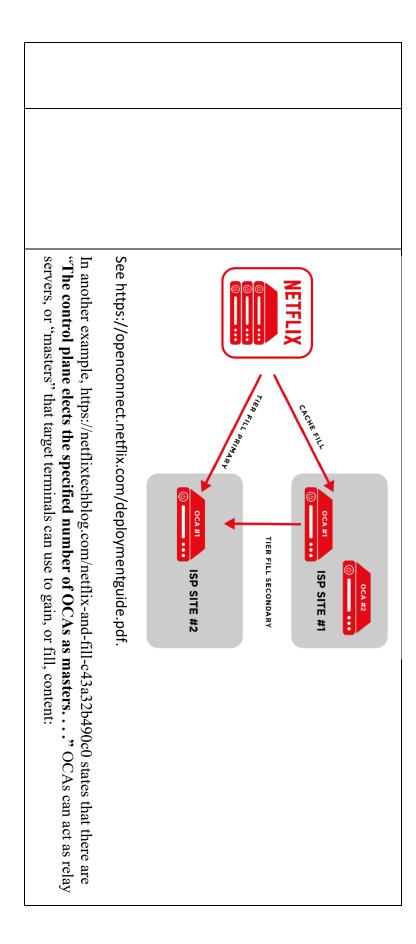
the health monitoring and OCA target terminal selection process described below in Sec. [1e].

The "target terminal selected form said plurality of terminals" language of the claim is infringed by

A "first server" is identified above in Sec. [1a] as a server within Netflix' data centers

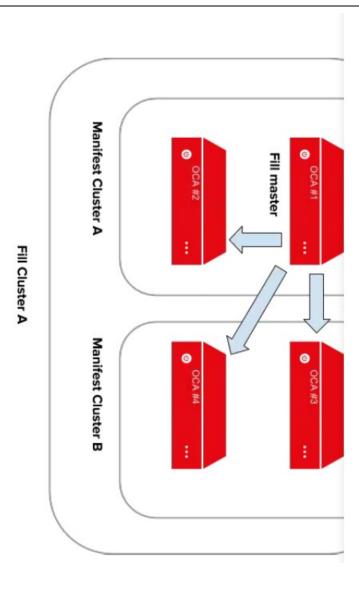
"target terminal" in this example that is served data from the relay server (OCA#1) retrieved from "relay server" to fill cached content to each other, from OCA#1 to OCA#2. OCA#2 becomes a peer filling content "cache fill" from the first server (data center) to OCA #1. OCA#1 then acts as a the first server (data center). See https://openconnect.netflix.com/deploymentguide.pdf. According to Netflix' network architecture, OCAs in a cluster and on the same subnet can attempt





The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to
3. Whether the OCA can go to S3, and how long it should wait before doing so
2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so
1. How many hops away an OCA can go to download content, and how long it should wait before doing so
A fill escalation policy defines:
• Fill escalation policies
• Fill master (number per fill cluster)
• BGP path attributes and physical location (latitude / longitude)
To calculate the least expensive fill source, we take into account network state and some configuration parameters for each OCA that are set by the Open Connect Operations team. For example:
It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to <i>all</i> of our OCAs, so we use a tiered approach. The goal is to ensure that the title is passed from one part of our network to another using the most efficient route possible.
• A calculated <i>route</i> cost — Described in the next section.
• Fill health — Can the fill source take on additional fill traffic?
• Title (content) availability — Does the fill source have the requested title stored:

should be storing and serving to members. If there is a delta between the list of titles during its configured fill window, that includes a list of the new or updated titles in the manifest and what they are currently storing, each OCA will send a request, requesting (among other things) a manifest file that contains the list of titles they plane. OCAs communicate at regular intervals with the control plane services, popularity, etc. All of this information is aggregated and stored in the AWS control OCAs do not store any information about other OCAs in the network, title Fill Source Manifests



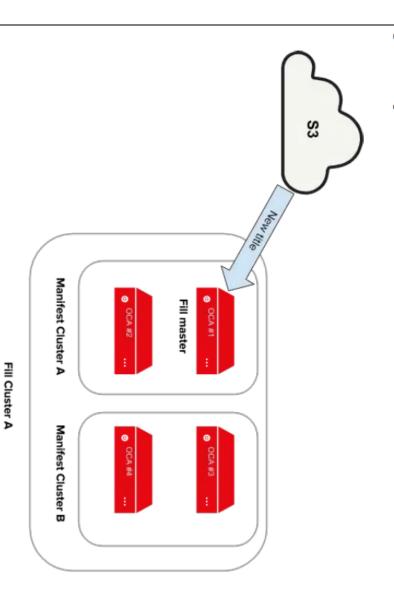
serve this new title have it stored. time zones — until enough of the OCAs in our global network that need to be able to time zone ends, and the fill pattern continues as the fill window moves across other that ensures we don't interrupt any live sessions. needed, they are put into a delete manifest and then deleted after a period of time during the fill window. If there are titles being stored on an OCA that are no longer status, other OCAs can then fill from them, and so on. This process continues As the sun moves west and more members begin streaming, the fill window in this When the second tier of OCAs complete their download, they report back their



2. Tier fill: Available OCAs outside the manifest cluster configuration

- Cache fill: Direct download from S3

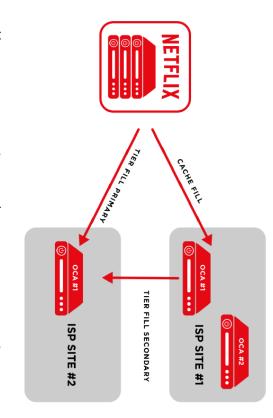
given the option to fill from the fill master. communicate with the control plane to request a fill source for this title, they are control plane that it now has the title stored. The next time the other OCAs After the fill master OCA has completed its S3 download, it reports back to the Example Scenario



[19c] a c c c c c c c c c c c c c c c c c c	a non-transitory computer usable medium having computer readable program code means embodied in said non-transitory medium, said computer readable program code program code means including:	It is on information and belief that Netflix Open Connect system includes the requisite software, or computer program, to carry out the functionality of Netflix's Open Connect system.
[19d] c	computer readable program code for causing said network server to manage selective retrieval of data from said first server by at least one target terminal selected from said plurality of terminals; and wherein said network server to monitor response times of terminals in the network and in which terminals	A "first server" or S3 is identified above in Sec. [1a] as a server within Netflix' data centers, and at least two of the OCAs are adapted to act as relay servers for serving data retrieved from said first server to at least one OCA terminal.  According to Netflix' network architecture, OCAs in a cluster and on the same subnet can attempt peer filling content "cache fill" from the first server to OCA #1. OCA#1 then acts as a "relay server" to fill cached content to each other, from OCA#1 to OCA#2. OCA#2 becomes a "target terminal" in this example that is served data from the relay server (OCA#1) retrieved from the first server (data center). See https://openconnect.netflix.com/deploymentguide.pdf.

which then acts as a "relay server" to fill a target terminal (OCA#2) using "tier filling." In another example, Netflix's first server (data center) can cache fill content to a terminal OCA#1,

cluster or subnet) will consider filling from each other rather than via a regular cache fill." considered. OCAs that can see each other's IP address in their BGP feed (but are not in the same "Peer filling is the most efficient method, however as a second-best option tier filling is



See https://openconnect.netflix.com/deploymentguide.pdf.

servers, or "masters" that target terminals can use to gain, or fill, content: "The control plane elects the specified number of OCAs as masters...." OCAs can act as relay In another example, https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states that there are

<ul> <li>Title (content) availability — Does the fill source have the requested title stored:</li> <li>Fill health — Can the fill source take on additional fill traffic?</li> </ul>
Calculating the Least Expensive Fill Source  It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to <i>all</i> of our OCAs, so we use a tiered approach. The goal is to ensure that the title is passed from one part of our network to another using the most efficient route possible.
To calculate the least expensive fill source, we take into account network state and some configuration parameters for each OCA that are set by the Open Connect Operations team. For example:
• BGP path attributes and physical location (latitude / longitude)
• Fill master (number per fill cluster)
• Fill escalation policies
A fill escalation policy defines:
1. How many hops away an OCA can go to download content, and how long it should wait before doing so
2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so
3. Whether the OCA can go to S3, and how long it should wait before doing so
The control plane elects the specified number of OCAs as masters for a given title asset. The fill escalation policies that are applied to masters typically allow them to  The main server on

relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a>, Netflix terminals are selected to act as relay servers for a particular data transfers on the basis of their CCS is further adapted to monitor response times of terminals in the network and in which

The response from the control plane in AWS is a ranked list of potential download locations, a.k.a fill sources, for each title. The determination of the list takes into consideration several high-level

Title (content) availability — Does the fill source have the requested title stored?

Fill health — Can the fill source take on additional fill traffic?

A calculated route cost — Described in the next section

A fill escalation policy defines:

- How many hops away an OCA can go to download content, and how long it should wait before doing so
- above), and how long it should wait before doing so Whether the OCA can go to the entire Open Connect network (beyond the hops defined

2

Whether the OCA can go to S3, and how long it should wait before doing so

(Emphasis added.)

should wait for a response before escalation to the next download locations. includes OCAs, and uses a fill escalation policy based upon response times to determine if the performance of the download location can take on additional traffic or not and how long the OCA The CCS server monitors the "fill health" of each potential download locations, or fill sources that

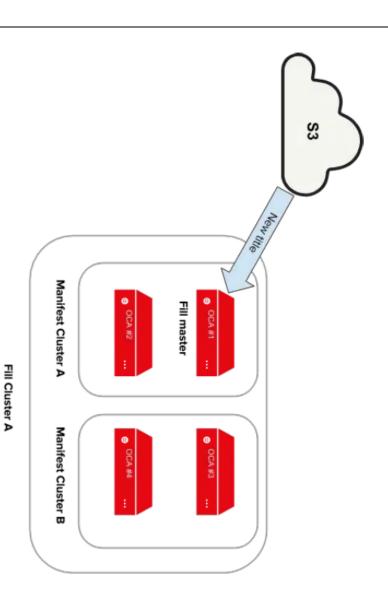
serve this new title have it stored.	time zones — until enough of the OCAs in our global network that need to be able to	time zone ends, and the fill pattern continues as the fill window moves across other	As the sun moves west and more members begin streaming, the fill window in this	that ensures we don't interrupt any live sessions.	needed, they are put into a delete manifest and then deleted after a period of time	during the fill window. If there are titles being stored on an OCA that are no longer	status, other OCAs can then fill from them, and so on. This process continues	When the second tier of OCAs complete their download, they report back their

# 2. Tier fill: Available OCAs outside the manifest cluster configuration 1. Peer fill: Available OCAs within the same manifest cluster or the same subnet

# 3. Cache fill: Direct download from S3

communicate with the control plane to request a fill source for this title, they are control plane that it now has the title stored. The next time the other OCAs After the fill master OCA has completed its S3 download, it reports back to the Example Scenario

given the option to fill from the fill master.



information in the network information stored performance the terminal terminal based on server/CCS server to a terminal/OCA to download content data in the manifest that needs to be direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix substantially the same. The claim states that the main server is adapted to send transport requests a terminal (OCA) with instructions to download or "fill" content. The way this is accomplished is This desired manifest and emergency manifest and the download location data on the CCS the network. The results are substantially the same - to transmit commands from the main posts a desired manifest or emergency manifest to the CCS server that is intended for each OCA in terminal." The functions are substantially the same - to send a request from a main server (CCS) to equivalent the "server is adapted to send transport requests direct to at least one first target are

terminals as a download location for a given title asset. The control plane elects the specified number of OCAs as masters for a given title asset. https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 Thus, the CCS server selects OCA

the desired manifest that is needed by an OCA can be downloaded manifest. OCAs then then query the CCS for location information files that list where each title on regularly from the CCS by each OCA terminal, which traverses the desired manifest or emergency hours download window, or urgently if there is an emergent need. these manifests are fetched by that indicate what titles need to be transferred to a specific OCA terminal, either during an off-peak Netflix to the CCS (Control Content Service) server in AWS. These manifests are data structures A "desired manifest" and when needed an "emergency manifest" are transport requests posted by

second OCA can be selected based on their relative performance: https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one

master OCAs: server, for a download location and provides other download locations of a second and additional The CCS server information location list provides the address of a first server, called the "S3"

all of our OCAs, so we use a tiered approach. . . . A fill escalation policy defines: It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to

wait before doing so	1. How many hops away an OCA can go to download content, and how long it should

- above), and how long it should wait before doing so 2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined
- 3. Whether the OCA can go to S3, and how long it should wait before doing sc

in order to grab that content and then share it locally with non-masters escalation policies that are applied to masters typically allow them to reach farther with less delay The control plane elects the specified number of OCAs as masters for a given title asset. The fill

is based on performance of the OCAs, to determine if that OCA will be selected as a download location or not. As stated in [1e], the CCS monitors the "fill health" and performance of download locations, which

one first target terminal over a network, and on the Open Connect Network, Netflix posts a desired main server is adapted to send transport requests that include download locations direct to at least content from specific master terminal/OCA addresses and includes an address of at least one second a request from a main server (CCS) to a terminal (OCA) with instructions to download or "fill" server to be relayed by the first target terminal." The functions are substantially the same - to send manifest that needs to be retrieved from various addresses commands from the main server/CCS server to a terminal/OCA to download content data in the for each OCA to read on a regular basis. The results are substantially the same - to transmit manifest or emergency manifest to the CCS server along with download locations that are intended terminal/OCA. The way this is accomplished is substantially the same. The claim states that the target terminal, the addresses of at least one second target terminal to which the data from the first to be retrieved, the address of the first server from which the data is to be requested by the first terminal) OCA on the CCS are equivalent the "each such transport request includes details of data files and fill policy for master OCAs and a second (target terminal) OCA to fill from a (first This desired manifest and emergency manifest along with the downloaded location information

and in which terminals are selected to act as relay servers for a particular data transfers on the basis The main server on CCS is further adapted to monitor response times of terminals in the network

OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle. The OCAs work in a network to distribute updates	a computer readable program code means for	[19f]
The CCS server monitors the "fill health" of each potential download locations, or fill sources that includes OCAs, and uses a fill escalation policy based upon response times to determine if the performance of the download location can take on additional traffic or not and how long the OCA should wait for a response before escalation to the next download locations. Thus, the manifest is assembled by the CCS server and download locations are assembled based on OCA performance information, and the manifest is sent to a given OCA. Therefore, the transport request is sent on a basis of said terminal performance information.		
(Emphasis added.)		
<ul> <li>16. How many hops away an OCA can go to download content, and how long it should wait before doing so</li> <li>17. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so</li> <li>18. Whether the OCA can go to S3, and how long it should wait before doing so</li> </ul>		
A fill escalation policy defines:		
A calculated route cost — Described in the next section.		
Fill health — Can the fill source take on additional fill traffic?		
Title (content) availability — Does the fill source have the requested title stored?		
The response from the control plane in AWS is a ranked list of potential download locations, a.k.a. fill sources, for each title. The determination of the list takes into consideration several high-level factors:		
of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a> . Netflix wrote:		

relay server; and

request is to act as modified transport further target addresses of further includes the modified in the transport terminals identified target terminals request to selected modified transport and to transmit the other relay servers server or from requests received to modify transport the recipient of the terminals for which transport request request, wherein from a set of target from said network

> among each other and to include further OCAs to which updates and content can be sent. See Open Connect Overview, p. 5; Fill Patterns, pp. 1-3.

network terminal

transport request to selected target terminals that includes addresses of further target terminals requests received from the main server or from other relay servers and transmit the modified Netflix' OCA that are adapted to act as relay servers (see 1d above) are adapted to modify transport

of these master or target terminals in the desired manifest, which is loaded by an OCA terminal in discriminator; 4) the geographically closest appliance. The CCS server includes the URL addresses the to transfer titles from: 1) the terminal appliance that receives the most-specific route; 2) the The CCS server will to order OCA terminals to peer or tier fill from using OCAs selected by the its memory or hard drive space in order to select an OCA for downloading titles from terminal appliance that receives the route to the client's netblock with the lowest multi-exit terminal appliance that receives the route to the client's netblock with the shortest AS path; 3) the CCS server. The CCS server uses Appliance Section Criteria to select OCA terminals as targets in

locations of master OCAs for each individual title needed by an OCA to fill its delta: on the delta. The CCS responds, as stated in [1f] with a list of URLs that are downloadable from the actual manifest are termed the "delta" or difference between actual and desired manifest. "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing The OCA terminal will then query the CCS terminal for a list of download locations for each title After receiving the desired manifest, an OCA terminal will compare the manifest with its own

https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 potential download locations, a.k.a. fill sources, for each title." (Emphasis added.) See updated titles that it needs. The response from the control plane in AWS is a ranked list of If there is a **delta** between the list of titles in the manifest and what they are currently storing, each things) a manifest file that contains the list of titles they should be storing and serving to members OCA will send a request, during its configured fill window, that includes a list of the new or "OCAs communicate at regular intervals with the control plane services, requesting (among other

and then requesting a delta-listed title from the list of a master OCAs, is equivalent to a modified The action of an OCA requesting download locations (master OCAs) for its delta list from the CCS

desired manifest) to request a title or titles from further target terminals, or master OCAs. transport request. The OCA is using a modified list of titles (delta or missing titles list from the

transport request to selected target terminals from a set of target terminals identified in the transport request, wherein the modified transport request further includes addresses of further target terminals requests received from the main server or from other relay servers and to transmit the modified This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport for which the recipient of the modified transport request is to act as relay server."

original transport request (desired manifest) is transmitted to selected relay terminals (single or substantially similar: further relay terminal addresses are sent to the OCAs, a modification of the of URL locations to download the modified list titles of its delta list, after which the delta list is another OCA, an "actual manifest" is sent to the CCS server which responds to the OCA with a list of what it is supposed to download. Instead of the terminal, or OCA, transmitting the delta list to substantially the same. A delta is a modification of the manifest list, or in other words a subset list with URLs of those relay servers or URLs of master OCAs. The way the actions are performed are relay servers or OCAs a modification of the original transport request, or desired manifest, along multiple delta titles to selected OCAs) are sent to addresses (URLs) of further selected relay servers then sent to another OCA via a series of modified transfer requests. The result of these actions are The function of the two actions are substantially the same, which is to transmit to other selected (OCAs) in the form of a request for one or more titles from one or more master OCAs.

clustered or if they are in the same subnet, will attempt to peer or tier fill from each other. https://openconnect.zendesk.com/hc/en-us/articles/360035618071-Fill-patterns Alternatively and equivalently, Netflix documentation discloses that OCA terminals, if they are

save this information. In general, appliances determine where to receive fill using selection criteria geographically closest terminal. The OCA terminal includes the URL addresses of these terminals that receives the route to the client's netblock with the lowest multi-exit discriminator; 4) the titles from: 1) the terminal appliance that receives the most-specific route; 2) the appliance terminal that receives the route to the client's netblock with the shortest AS path; 3) the appliance terminal Criteria as the CCS server uses to select OCA terminals as targets in the subnet or cluster to transfer that is used by Netflix client devices. The OCA terminals then use a similar Appliance Section OCA terminals in a subnet or cluster broadcast their IP and physical locations to one another and

T				
[19g]				
wherein data to be retrieved by said target terminals are divided into a series of packets for transmission to said target				
Netflix' OCAs are adapted to communicate with the main server which is hosted by AWS. The OCA's "Report their status to the Open Connect control panel services in Amazon Web Services": The Open Connect network can make parallelized cache fill transfers. As such, it must be transferring "packets" of the file. There is evidence that each file is downloaded for cache fill in packet sizes up to 16 kilobytes.	The function of the two actions are substantially the same, which is to transmit to other selected relay servers or OCAs a modified list of the original transport request or manifest. The way the actions are performed are substantially the same. A modified desired manifest list is a subset list called the "delta." A relay server (OCA) creates a modified transport request (delta list) of its missing titles and transports this list to another relay server (OCA sends requests to other selected OCAs for the titles on the delta list in a series of requests) and includes addresses of the other relay servers (URLs of the selected OCA terminals in the request as an address to the selected OCAs). The result of these actions are substantially similar: a modification of the original transport request (desired manifest) is transmitted to selected relay terminals (selected OCAs) with addresses (URLs) of the selected relay servers (OCAs).	This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport requests received from the main server or from other relay servers and to transmit the modified transport request to selected target terminals from a set of target terminals identified in the transport request, wherein the modified transport request further includes addresses of further target terminals for which the recipient of the modified transport request is to act as relay server."	After receiving the desired manifest, an OCA terminal will compare the manifest with its own "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing from the actual manifest are termed the "delta" or difference between actual and desired manifest. After selecting an OCA master using the selection criteria, the OCA terminal will transmit and request the delta list items to the selected OCA(s) in the subnet or cluster in the form of download requests for each title using the URL of the target OCA.	in its memory or hard drive space in order to select an OCA for downloading titles from. See <i>Fill Patterns</i> , pp. 1-4.

					packet of a series routed thereto.	directly with said main server to acknowledge receipt of the last	terminals and each of said terminals are adapted to communicate
https://openconnect.zendesk.com/hc/en-us/articles/360035533071#routi	<ul> <li>Traffic from OCA: Allow all destination addresses and ports.</li> <li>Traffic to OCA: Allow TCP 22, 53, 80, 179, 443, UDP 53 and 123 (source and destination), ICMP types 0, 3, 8, 11, and all ICMPv6 from any public IP/port. Allow all return traffic from any appliance-initiated connection (TCP established).</li> </ul>	Netflix requires network traffic to OCA be in TCP protocol:	Comer, Douglas E. (2006). Internetworking with TCP/IP: Principles, Protocols, and Architecture. Vol. 1 (5th ed.). Prentice Hall. ISBN 978-0-13-187671-2	TCP is a reliable byte stream delivery service which guarantees that all bytes received will be identical and in the same order as those sent. Since packet transfer by many networks is not reliable, TCP achieves this using a technique known as positive acknowledgement with retransmission. This requires the receiver to respond with an acknowledgement message as it receives the data. The sender keeps a record of each packet it sends and maintains a timer from when the packet was sent. The sender re-transmits a packet if the timer expires before receiving the acknowledgement. The timer is needed in case a packet gets lost or corrupted.	TCP data communication protocol, which most of the Internet included Netflix uses, requires an acknowledgement of packets when content is transmitted:	There is also evidence that the cache fill transfer uses TCP/IP protocols for packetized data transfer.	Additionally, the last packet of a download for a cache fill is associated with a notice to the CCS that the download has been completed. The system can also use parallel processes to perform cache transfer using TCP/IP protocols. This is only possible if the CCS is working on a different section of the file at the same time, which is packetized file transfer.

37

# Glossary of terms

in a local server of files to be delivered to consumers, stored originated elsewhere. Thus for CDNs, a copy

## **CDN (Content Delivery Network)**

content over the internet enable the efficient and reliable distribution of A distributed system of servers, designed to

### CP (content provider)

content (rather than - say - e-commerce) An internet business whose focus is delivering to consumers. Netflix, CNN and YouTube

# One step in a packet of data's journey

횽

Fixed broadband

transmits content A computer that stores and

### TCP/IP

ISP (Internet Service Provider)

across multiple servers

A company providing internet connectivity

data transmission that underpin the internet protocol. The two foundational standards for Transmission control protocol and internet

IXP (Internet Exchange Point)

provide fixed broadband, mobile data or both to consumers (consumers or businesses). May

to exchange traffic, avoiding the need A location where many networks meet

for multiple bilateral connections

Compound Annual Growth Rate

sent and reaching its destination

OTT (Over The Top)

The lag between a packet of data being

A temporary local copy of information that

### Packet loss

broadband networks)

(since they are not provided by telcos operating YouTube, Facebook and Netflix are examples network without being integrated with it. Describes services delivered over another

handle, it discards a certain amount of data When a router is sent more data than it can

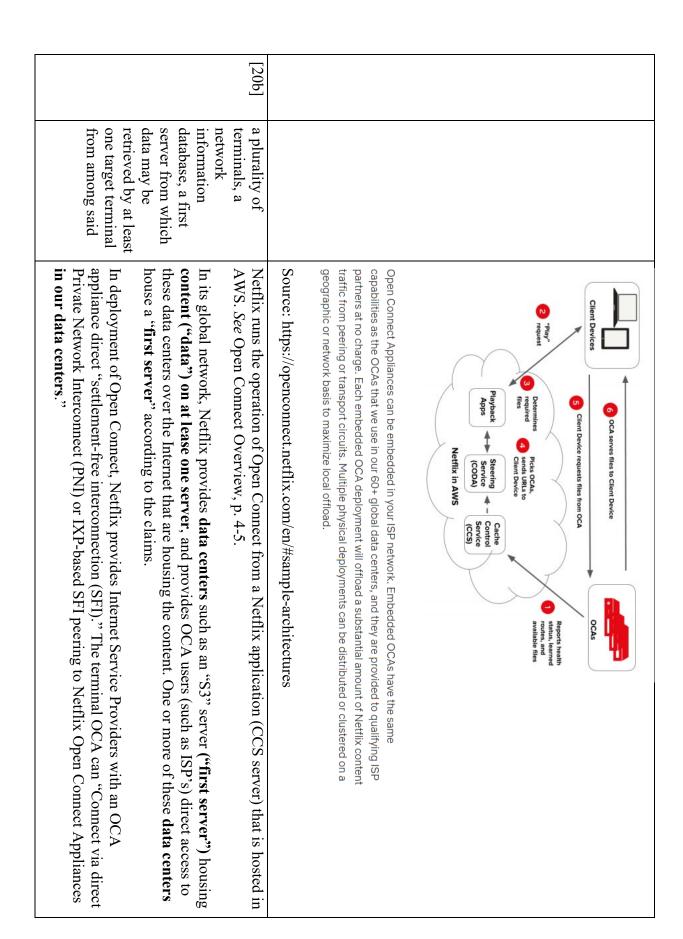
from the source server

data in question will then be requested again This is known as packet loss. Typically the

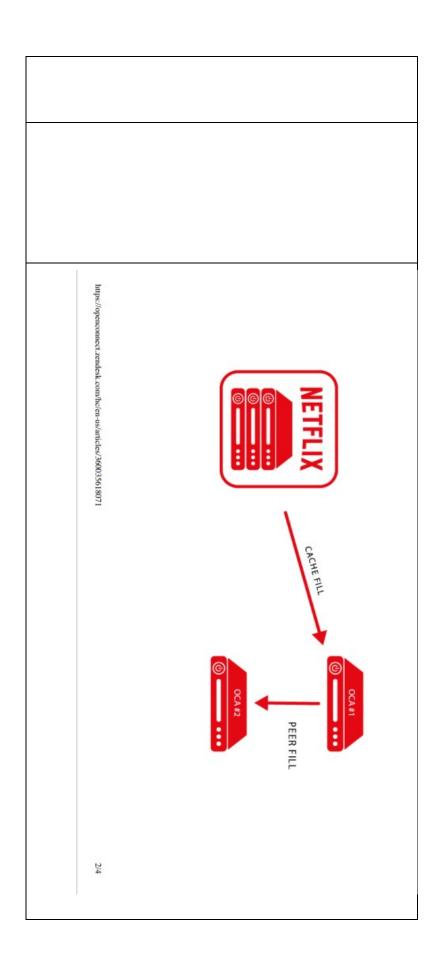
of data and sends them onwards down the Server appropriate link A switch on the internet, that receives packets

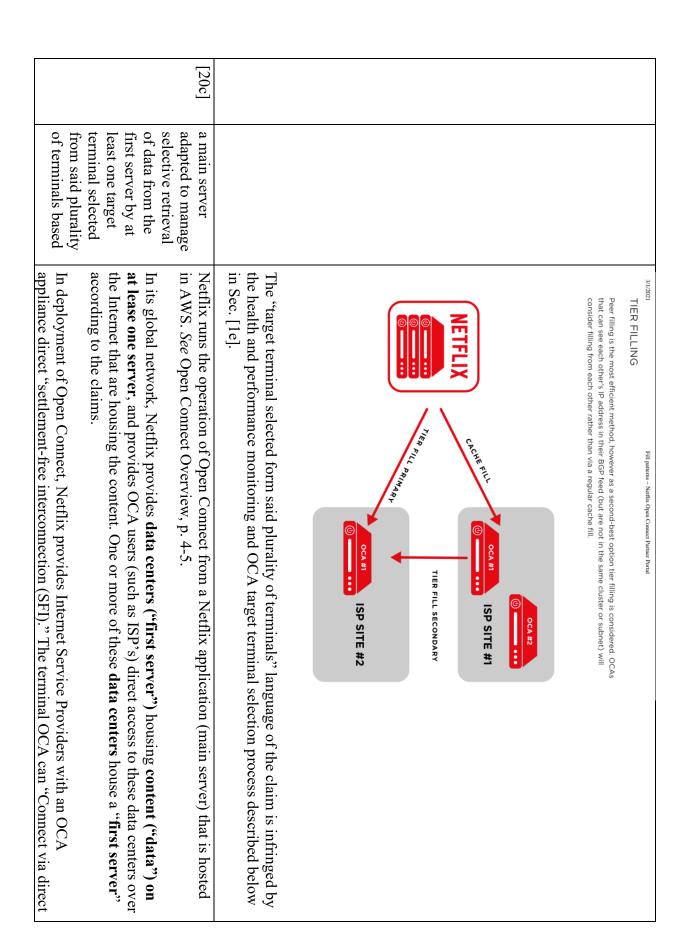
### 131

and display mechanism.	network, the data communication network including:	
Connect control plane (CCS server) to control fill behavior (adding new files to OCAs nightly) and to compute and/or store data. <i>See id.</i> p. 3-4. Accordingly, OCAs include both an input mechanism	data communication	
directly inside ISP networks. Netflix provides the server hardware. The OCAs report to a Open	to operate as a relay server in a	
The building blocks of Open Connect are our suite of purpose-built server appliances, called	for enabling a network terminal	
world-wide.	program product	
Netflix uses a system called Open Connect to deliver Netflix TV shows and movies to members	A computer	[20a]
In another example, "After the fill master OCA has completed its S3 download, it reports back to the control plane that it now has the title stored." See https://netflixtechblog.com/netflix-and-fill-c43a32b490c0		
Netflix uses TCP for internet streaming to send packets of data for video. Additionally, Netflix specifically looks at the number of TCP connections to determine internet speeds in accordance with testing of OCAs.		
Where does Netflix use TCP?		
"A cooperative approach to content delivery," Netflix (2021), 37.		



				plurality of terminals; and
OCAs in a cluster and on the same subnet can attempt peer filling from each other. There is also Tier filling where if in different ISP sites. <a href="deploymentguide.pdf">deploymentguide.pdf</a> (netflix.com)	The following diagram also illustrates access from a target terminal (OCA $\#1$ , OCA $\#2$ ) to a Netflix first server in "our data centers."	Welcome to Open Connect, p. 3. Dkt 39 at p. 48.	ISPs who do not currently participate in pubic peering might want to consider that a single IX port can support multiple peering sessions, providing <b>direct access to various content</b> , cloud, and network providers.	"Netflix has the ability to interconnect at a number of global data center facilities and public Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any network at IXP locations where we are mutually present and we consider private interconnection as appropriate."





stored in the

on terminal

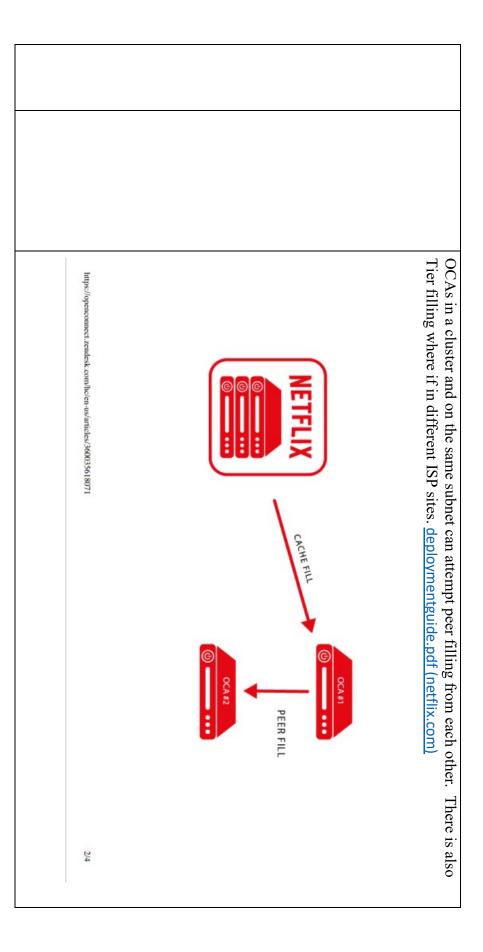
appropriate." network at IXP locations where we are mutually present and we consider private interconnection as Internet Exchange fabrics as listed on our Peering Locations page. We openly peer with any "Netflix has the ability to interconnect at a number of global data center facilities and public

network providers. can support multiple peering sessions, providing direct access to various content, cloud, and ISPs who do not currently participate in pubic peering might want to consider that a single IX port

Welcome to Open Connect, p. 3. Dkt 39 at p. 48

as relay servers for are selected to act which terminals network and in terminals in the response times of server to monitor wherein the main database, and information

The following diagram also illustrates access from a target terminal (OCA #1, OCA #2) to a Netflix first server in "our data centers."



### consider filling from each other rather than via a regular cache fill. that can see each other's IP address in their BGP feed (but are not in the same cluster or subnet) will Peer filling is the most efficient method, however as a second-best option tier filling is considered. OCAs TIER FILLING OCA #1 TIER FILL SECONDARY **ISP SITE #2 ISP SITE #1**

the health monitoring and OCA target terminal selection process described below in Sec. [1e]. The "target terminal selected form said plurality of terminals" language of the claim is infringed by

servers, or "masters" that target terminals can use to gain, or fill, content: "The control plane elects the specified number of OCAs as masters..." OCAs can act as relay In another example, https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states that there are

• Fill health — Can the fill source take on additional fill traffic?	<ul> <li>Title (content) availability — Does the fill source have the requested title stored:</li> </ul>

# A calculated route cost — Described in the next section.

# Calculating the Least Expensive Fill Source

route possible. the title is passed from one part of our network to another using the most efficient from S3 to all of our OCAs, so we use a tiered approach. The goal is to ensure that It would be inefficient, in terms of both time and cost, to distribute a title directly

Operations team. For example: some configuration parameters for each OCA that are set by the Open Connect To calculate the least expensive fill source, we take into account network state and

- BGP path attributes and physical location (latitude / longitude)
- Fill master (number per fill cluster)
- Fill escalation policies

A fill escalation policy defines:

- 1. How many hops away an OCA can go to download content, and how long it should wait before doing so
- 2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so
- 3. Whether the OCA can go to S3, and how long it should wait before doing so

asset. The fill escalation policies that are applied to masters typically allow them to The control plane elects the specified number of OCAs as masters for a given title

receiving and responding to transport requests sent to said network terminal, each such transport request including details of data to be	[20e] computer readable program code for causing said network terminal to act as relay server for serving data retrieved from said first server to at least one target terminal by  All OCA deployments are constantly monitored to ensure reliability and efficiency. Netflix makes use of non-peak bandwidth to download the vast majority of content updates to the OCAs in network terminal by use of non-peak bandwidth to download the vast majority of content updates from network terminal by use of non-peak bandwidth to download the vast majority of content updates to the OCAs in network terminal by use of non-peak bandwidth to download the vast majority of content updates to the OCAs in network terminal by use of non-peak bandwidth to download the vast majority of content updates to the OCAs in network terminal by and efficiency. Netflix makes use of non-peak bandwidth to download the vast majority of content updates to the OCAs in network terminal by acknowled updates from network during significant usage of internet "backbone" capacity during the update cycle.  Server for serving A "desired manifest" and when needed an "emergency manifest" are posted by Netflix to the CCS as a specific OCA terminal, either during an off-peak hours download window, or urgently if there is an emergent need.	[20d] a non-transitory computer usable medium having computer readable program code means embodied in said non-transitory medium, said computer readable program code means including:
mergency manifest" are posted by Netflix to the CO hese manifests are data structures that indicate what terminal, either during an off-peak hours downloaneed. Equivalently, these manifests are fetched by ral, which traverses the desired manifest or emergen abox for manifests sent directly to the OCAs, where yularly for manifests and download the manifests to	red to ensure reliability and efficiency. Netflix make he vast majority of content updates to the OCAs indows. OCAs can also download updates from of internet "backbone" capacity during the updates mergency manifest" are posted by Netflix to the Conctures that indicate what titles need to be transferre f-peak hours download window, or urgently if there	Netflix Open Connect system includes the requisite the functionality of Netflix's Open Connect system

the data is to be

information in the network performance stored further target performance of a of a relative and an indication network terminal relayed by the which the data target terminal to network terminal, requested by the the terminal terminal based on first server is to be retrieved from the the addresses of at least one second

server/CCS server to a terminal/OCA to download content data in the manifest that needs to be the network. The results are substantially the same - to transmit commands from the main posts a desired manifest or emergency manifest to the CCS server that is intended for each OCA in direct to at least one first target terminal over a network, and on the Open Connect Network, Netflix substantially the same. The claim states that the main server is adapted to send transport requests a terminal (OCA) with instructions to download or "fill" content. The way this is accomplished is retrieved. terminal." The functions are substantially the same - to send a request from a main server (CCS) to

the desired manifest that is needed by an OCA can be downloaded regularly from the CCS by each OCA terminal, which traverses the desired manifest or emergency that indicate what titles need to be transferred to a specific OCA terminal, either during an off-peak Netflix to the CCS (Control Content Service) server in AWS. These manifests are data structures manifest. OCAs then then query the CCS for location information files that list where each title on hours download window, or urgently if there is an emergent need. these manifests are fetched by A "desired manifest" and when needed an "emergency manifest" are transport requests posted by

second OCA can be selected based on their relative performance: https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 states a master OCA and at least one

server, for a download location and provides other download locations of a second and additional master OCAs: The CCS server information location list provides the address of a first server, called the "S3"

all of our OCAs, so we use a tiered approach. . . . A fill escalation policy defines: It would be inefficient, in terms of both time and cost, to distribute a title directly from S3 to

- wait before doing so 1. How many hops away an OCA can go to download content, and how long it should
- above), and how long it should wait before doing sc 2. Whether the OCA can go to the entire Open Connect network (beyond the hops defined

# Whether the OCA can go to S3, and how long it should wait before doing so

in order to grab that content and then share it locally with non-masters escalation policies that are applied to masters typically allow them to reach farther with less delay The control plane elects the specified number of OCAs as masters for a given title asset. The fill

is based on performance of the OCAs, to determine if that OCA will be selected as a download location or not. As stated in [1e], the CCS monitors the "fill health" and performance of download locations, which

manifest or emergency manifest to the CCS server along with download locations that are intended one first target terminal over a network, and on the Open Connect Network, Netflix posts a desired content from specific master terminal/OCA addresses and includes an address of at least one second a request from a main server (CCS) to a terminal (OCA) with instructions to download or "fill" server to be relayed by the first target terminal." The functions are substantially the same - to send main server is adapted to send transport requests that include download locations direct to at least terminal/OCA. The way this is accomplished is substantially the same. The claim states that the target terminal, the addresses of at least one second target terminal to which the data from the first to be retrieved, the address of the first server from which the data is to be requested by the first terminal) OCA on the CCS are equivalent the "each such transport request includes details of data files and fill policy for master OCAs and a second (target terminal) OCA to fill from a (first This desired manifest and emergency manifest along with the downloaded location information

terminals as a download location for a given title asset. The control plane elects the specified number of OCAs as masters for a given title asset https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 Thus, the CCS server selects OCA

manifest that needs to be retrieved from various addresses

commands from the main server/CCS server to a terminal/OCA to download content data in the

for each OCA to read on a regular basis. The results are substantially the same - to transmit

and in which terminals are selected to act as relay servers for a particular data transfers on the basis The main server on CCS is further adapted to monitor response times of terminals in the network

OCAs can also download updates from each other – minimizing significant usage of internet "backbone" capacity during the update cycle. The OCAs work in a network to distribute updates	said computer readable program code for causing	[20f]
The CCS server monitors the "fill health" of each potential download locations, or fill sources that includes OCAs, and uses a fill escalation policy based upon response times to determine if the performance of the download location can take on additional traffic or not and how long the OCA should wait for a response before escalation to the next download locations. Thus, the manifest is assembled by the CCS server and download locations are assembled based on OCA performance information, and the manifest is sent to a given OCA. Therefore, the transport request is sent on a basis of said terminal performance information.		
(Emphasis added.)		
<ul> <li>19. How many hops away an OCA can go to download content, and how long it should wait before doing so</li> <li>20. Whether the OCA can go to the entire Open Connect network (beyond the hops defined above), and how long it should wait before doing so</li> <li>21. Whether the OCA can go to S3, and how long it should wait before doing so</li> </ul>		
A fill escalation policy defines:		
A calculated route cost — Described in the next section.		
Fill health — Can the fill source take on additional fill traffic?		
Title (content) availability — Does the fill source have the requested title stored?		
The response from the control plane in AWS is a ranked list of potential download locations, a.k.a. fill sources, for each title. The determination of the list takes into consideration several high-level factors:		
of their relative response times. At <a href="https://netflixtechblog.com/netflix-and-fill-c43a32b490c0">https://netflixtechblog.com/netflix-and-fill-c43a32b490c0</a> , Netflix wrote:		

relay server; and

request is to act as modified transport further target addresses of further includes the modified in the transport terminals identified target terminals request to selected modified transport transmit the servers and to main server or received from the transport requests said network the recipient of the terminals for which transport request request, wherein from a set of target from other relay terminal to modify

> Connect Overview, p. 5; Fill Patterns, pp. 1-3 among each other and to include further OCAs to which updates and content can be sent. See Open

transport request to selected target terminals that includes addresses of further target terminals requests received from the main server or from other relay servers and transmit the modified Netflix' OCA that are adapted to act as relay servers (see 1d above) are adapted to modify transport

of these master or target terminals in the desired manifest, which is loaded by an OCA terminal in discriminator; 4) the geographically closest appliance. The CCS server includes the URL addresses the to transfer titles from: 1) the terminal appliance that receives the most-specific route; 2) the its memory or hard drive space in order to select an OCA for downloading titles from terminal appliance that receives the route to the client's netblock with the lowest multi-exit terminal appliance that receives the route to the client's netblock with the shortest AS path; 3) the CCS server. The CCS server uses Appliance Section Criteria to select OCA terminals as targets in The CCS server will to order OCA terminals to peer or tier fill from using OCAs selected by the

locations of master OCAs for each individual title needed by an OCA to fill its delta: on the delta. The CCS responds, as stated in [1f] with a list of URLs that are downloadable from the actual manifest are termed the "delta" or difference between actual and desired manifest. "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing The OCA terminal will then query the CCS terminal for a list of download locations for each title After receiving the desired manifest, an OCA terminal will compare the manifest with its own

https://netflixtechblog.com/netflix-and-fill-c43a32b490c0 potential download locations, a.k.a. fill sources, for each title." (Emphasis added.) See updated titles that it needs. The response from the control plane in AWS is a ranked list of If there is a **delta** between the list of titles in the manifest and what they are currently storing, each things) a manifest file that contains the list of titles they should be storing and serving to members OCA will send a request, during its configured fill window, that includes a list of the new or "OCAs communicate at regular intervals with the control plane services, requesting (among other

and then requesting a delta-listed title from the list of a master OCAs, is equivalent to a modified The action of an OCA requesting download locations (master OCAs) for its delta list from the CCS

desired manifest) to request a title or titles from further target terminals, or master OCAs. transport request. The OCA is using a modified list of titles (delta or missing titles list from the

transport request to selected target terminals from a set of target terminals identified in the transport request, wherein the modified transport request further includes addresses of further target terminals requests received from the main server or from other relay servers and to transmit the modified This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport for which the recipient of the modified transport request is to act as relay server."

original transport request (desired manifest) is transmitted to selected relay terminals (single or substantially similar: further relay terminal addresses are sent to the OCAs, a modification of the of URL locations to download the modified list titles of its delta list, after which the delta list is another OCA, an "actual manifest" is sent to the CCS server which responds to the OCA with a list of what it is supposed to download. Instead of the terminal, or OCA, transmitting the delta list to substantially the same. A delta is a modification of the manifest list, or in other words a subset list with URLs of those relay servers or URLs of master OCAs. The way the actions are performed are relay servers or OCAs a modification of the original transport request, or desired manifest, along multiple delta titles to selected OCAs) are sent to addresses (URLs) of further selected relay servers then sent to another OCA via a series of modified transfer requests. The result of these actions are The function of the two actions are substantially the same, which is to transmit to other selected (OCAs) in the form of a request for one or more titles from one or more master OCAs.

clustered or if they are in the same subnet, will attempt to peer or tier fill from each other. https://openconnect.zendesk.com/hc/en-us/articles/360035618071-Fill-patterns Alternatively and equivalently, Netflix documentation discloses that OCA terminals, if they are

save this information. In general, appliances determine where to receive fill using selection criteria geographically closest terminal. The OCA terminal includes the URL addresses of these terminals that receives the route to the client's netblock with the lowest multi-exit discriminator; 4) the titles from: 1) the terminal appliance that receives the most-specific route; 2) the appliance terminal that receives the route to the client's netblock with the shortest AS path; 3) the appliance terminal Criteria as the CCS server uses to select OCA terminals as targets in the subnet or cluster to transfer that is used by Netflix client devices. The OCA terminals then use a similar Appliance Section OCA terminals in a subnet or cluster broadcast their IP and physical locations to one another and

[20g]				
wherein data to be retrieved by said target terminals are divided into a series of packets for transmission to said target				
Netflix' OCAs are adapted to communicate with the main server which is hosted by AWS. The OCA's "Report their status to the Open Connect control panel services in Amazon Web Services": The Open Connect network can make parallelized cache fill transfers. As such, it must be transferring "packets" of the file. There is evidence that each file is downloaded for cache fill in parallel in packet sizes up to 16 kilobytes.	The function of the two actions are substantially the same, which is to transmit to other selected relay servers or OCAs a modified list of the original transport request or manifest. The way the actions are performed are substantially the same. A modified desired manifest list is a subset list called the "delta." A relay server (OCA) creates a modified transport request (delta list) of its missing titles and transports this list to another relay server (OCA sends requests to other selected OCAs for the titles on the delta list in a series of requests) and includes addresses of the other relay servers (URLs of the selected OCA terminals in the request as an address to the selected OCAs). The result of these actions are substantially similar: a modification of the original transport request (desired manifest) is transmitted to selected relay terminals (selected OCAs) with addresses (URLs) of the selected relay servers (OCAs).	This is equivalent to: "terminals adapted to act as relay servers are adapted to modify transport requests received from the main server or from other relay servers and to transmit the modified transport request to selected target terminals from a set of target terminals identified in the transport request, wherein the modified transport request further includes addresses of further target terminals for which the recipient of the modified transport request is to act as relay server."	After receiving the desired manifest, an OCA terminal will compare the manifest with its own "actual manifest", or the list of titles currently residing with the OCA terminal. The titles missing from the actual manifest are termed the "delta" or difference between actual and desired manifest. After selecting an OCA master using the selection criteria, the OCA terminal will transmit and request the delta list items to the selected OCA(s) in the subnet or cluster in the form of download requests for each title using the URL of the target OCA.	in its memory or hard drive space in order to select an OCA for downloading titles from. See <i>Fill Patterns</i> , pp. 1-4.

					packet of a ser routed thereto.	main server to acknowledge receipt of the l	communicate directly with	are adapted to	terminal of said t
					packet of a series routed thereto.	main server to acknowledge receipt of the last	communicate directly with said	ted to	terminals and each of said terminals
https://openconnect.zendesk.com/hc/en-us/articles/360035533071#routi	<ul> <li>Traffic from OCA: Allow all destination addresses and ports.</li> <li>Traffic to OCA: Allow TCP 22, 53, 80, 179, 443, UDP 53 and 123 (source and destination), ICMP types 0, 3, 8, 11, and all ICMPv6 from any public IP/port. Allow all return traffic from any appliance-initiated connection (TCP established).</li> </ul>	Netflix requires network traffic to OCA be in TCP protocol:	Comer, Douglas E. (2006). Internetworking with TCP/IP: Principles, Protocols, and Architecture. Vol. 1 (5th ed.). Prentice Hall. ISBN 978-0-13-187671-2	TCP is a reliable byte stream delivery service which guarantees that all bytes received will be identical and in the same order as those sent. Since packet transfer by many networks is not reliable, TCP achieves this using a technique known as positive acknowledgement with retransmission. This requires the receiver to respond with an acknowledgement message as it receives the data. The sender keeps a record of each packet it sends and maintains a timer from when the packet was sent. The sender re-transmits a packet if the timer expires before receiving the acknowledgement. The timer is needed in case a packet gets lost or corrupted.	TCP data communication protocol, which most of the Internet included Netflix uses, requires an acknowledgement of packets when content is transmitted:	There is also evidence that the cache fill transfer uses TCP/IP protocols for packetized data transfer.	of the file at the same time, which is packetized file transfer.	transfer using TCP/IP protocols. This is only possible if the CCS is working on a different section	Additionally, the last packet of a download for a cache fill is associated with a notice to the CCS that the download has been completed. The system can also use parallel processes to perform cache

37

# Glossary of terms

Compound Annual Growth Rate

A temporary local copy of information that in a local server of files to be delivered to consumers, stored originated elsewhere. Thus for CDNs, a copy

## **CDN (Content Delivery Network)**

content over the internet enable the efficient and reliable distribution of A distributed system of servers, designed to

### CP (content provider)

content (rather than - say - e-commerce) An internet business whose focus is delivering to consumers. Netflix, CNN and YouTube

# across multiple servers

One step in a packet of data's journey

Fixed broadband

of data and sends them onwards down the

A switch on the internet, that receives packets

appropriate link

provide fixed broadband, mobile data or both to consumers (consumers or businesses). May A company providing internet connectivity

IXP (Internet Exchange Point)

to exchange traffic, avoiding the need A location where many networks meet

for multiple bilateral connections

sent and reaching its destination The lag between a packet of data being

### OTT (Over The Top)

broadband networks) (since they are not provided by telcos operating YouTube, Facebook and Netflix are examples network without being integrated with it. Describes services delivered over another

### Packet loss

from the source server data in question will then be requested again This is known as packet loss. Typically the handle, it discards a certain amount of data When a router is sent more data than it can

transmits content A computer that stores and Server

### TCP/IP

ISP (Internet Service Provider)

data transmission that underpin the internet protocol. The two foundational standards for Transmission control protocol and internet

In another example, "After <b>the fill master OCA</b> has completed its S3 download, <b>it reports back to the control plane</b> that it now has the title stored." See https://netflixtechblog.com/netflix-and-fill-c43a32b490c0	Netflix uses TCP for internet streaming to send packets of data for video. Additionally, Netflix specifically looks at the number of TCP connections to determine internet speeds in accordance with testing of OCAs.	Where does Netflix use TCP?	"A cooperative approach to content delivery," Netflix (2021), 37.